Technical Report

AD 640 122

MECHANIZATION STUDY OF THE TECHNICAL LIBRARY, NAVAL ORDNANCE LABORATORY, WHITE OAK, MARYLAND

Submitted to

Defense Supply Agency
Defense Documentation Center
Cameron Station, Virginie

by

Booz, Allen Applied Research Inc.
4733 Bethesda Avenue
Bethesda, Maryland 20014

Under Contract No. DSA-7-15489

BAARINC Report No. 914-1-27

September 1966

BOOZ·ALLEN APPLIED RESEARCH INC.

WASHINGTON
CLEVELAND
CHICAGO
LOS ANGELES

ABSTRACT

The Library at the Naval Ordnance Laboratory (NOL) uses the IBM 7090 computer for information storage and retrieval of its technical reports collection. All technical reports are eventually entered on microfilm, with the present collection totaling about 150, 000. However, only documents received from 1958 to the present are available for computer searches. Plans call for an annual listing, by EAM techniques, of periodicals received at NOL. During fiscal 1966, punched card preparation will be completed. In addition, NOL and the Applied Physics Laboratory at Johns Hopkins are considering the possibility of cooperatively producing an annual listing of periodicals available in technical libraries in the Washington-Maryland area. There is also a study being conducted at present to determine the feasibility of a printed book catalog. The computer program has provided more exhaustive searches and more detailed subject analysis of reports and has eliminated the time lag in processing subject information.

TABLE OF CONTENTS

A P P E N D I C E S

A. INSTRUCTIONS FOR USE OF INPUT DATA WORKSHEET INCLUDING USE OF DESCRIPTOR CODES

B. EXAMPLE OF COMPLETED INPUT DATA WORKSHEET AND RESULTANT PUNCHED CARDS

C. SAMPLE SEARCH OUTPUTS

D. CATALOG CARDS

E. CHARGE CARDS

F. ACCESSIONS LIST

G. DEVELOPMENT OF DESCRIPTORS AND COMPUTER CODES

H. SAMPLES OF DESCRIPTORS AND COMPUTER CODES

I. SPECIAL RECORD

J. MASTER DICTIONARY RECORD

K. FORMING THE ZATO CODES

L. DOC ARRAY

M. PACKING THE DESCRIPTOR RANKS

N. DOCUMENT FILE RECORD
(ALSO BATCH ARRAY)

O. TITLE FILE RECORD
A AND B ARRAYS (STACKED TITLE TAPE)
DELET ARRAY

INDEX OF FIGURES

I. SUMMARY

The Library at the Naval Ordnance Laboratory (NOL) is utilizing the IBM 7090 computer for information storage and retrieval of its technical reports collection. The computer program has provided more exhaustive searches and more detailed subject analysis of reports, and it has eliminated the time lag in processing subject information. All technical reports are eventually entered on microfilm, with the present collection totaling approximately 150, 000. Only documents received from 1958 to the present are available for computer searches. In the past fiscal year, 19, 185 reports were added to the computer files, bringing the total to 66, 058. Computer searches for the same year number 2, 299, a decrease of 745 from the previous year.

The Naval Ordnance Laboratory, a field activity of the Bureau of Naval Weapons, Navy Department, is engaged in research and development in ordnance engineering. The Laboratory has 1, 100 scientific and engineering personnel and 2, 000 management craft, and clerical personnel. NOL's Library staff of 25 is organized into a Reference Acquisitions Branch and a Catalog Branch. It provides reference and bibliographic services on books, periodicals, and

technical reports to Laboratory research and engineering personnel and to the contractors who assist the Laboratory in accomplishing its mission. These users may request translations, abstracts, bibliographies, ready reference information, and specific items by telephone, memorandum, or in person. The Library staff provides assistance as necessary.

The Library book collection consists of 40,800 titles, with an annual growth of 700. Periodical subscriptions number 700 with a total of 1,100 copies; many of these are routed to NOL personnel, both administrative and technical. Punched cards are used for circulation of books. The borrower's payroll number is used as identification. An accessions list for books is typed monthly. Although provision exists for producing charge-out lists, this is not presently being done. Future plans call for an annual listing of periodicals. There is also a study being conducted at present to determine the feasibility of a printed book catalog.

II. MECHANIZATION

1. CHRONOLOGY

In early 1957, NOL's Librarian spent two weeks in California visiting Navy libraries in the Los Angeles area to study contemplated and operational computer information systems. From March to June of 1957, the Librarian and a NOL mathematician developed plans for mechanizing library functions.

By April 1958, 1,000 reports had been keypunched and were ready for testing. Also, programs were being written by two programmers.

Early in 1959, an IBM 704 was installed in the Mathematics Department, and the Library was allotted time on the computer for a library retrieval program. Descriptors and codes were developed from the 2,500 subject terms in the library catalog. In 1960, the information storage and retrieval program became operational.

In 1961 and 1962, revisions were made in the Library's information retrieval program due to the changeover from the IBM 704 to the IBM 7090. Conversion of the program was accomplished by the Mathematics Department; changes in the Library computer program at this time provided additional information for Library computer searches.

Also during these years, a Work Simplification Program continued and a concentrated effort was made to keep the Library staff informed of plans and developments in the computer programs. Each was carried out with group discussions, personal interviews, and circulation of pertinent literature. Also, a second Friden Programmatic Flexowriter and IBM 026 keypunch were acquired for use in library retrieval.

In 1963, an auxiliary program (Title Program) was developed to permit descriptive cataloging information available from the Flexowriter paper tape to be used as input to the Library Search Program. As a result of this program, a second printout for computer searches is produced, which now gives complete descriptive cataloging information for report identification of approximately 7,000 titles.

Through an informal survey of user reaction that was completed in 1964, it was found that users are satisfied with the computer searches about 85 percent of the time. The unsatisfied 15 percent covers requests for additional information from the computer and/or problems in retrieving the required information resulting from poor communication, searches which are too broad or too narrow in scope to satisfy the Library patron, or incorrect information furnished as to source, date of report, subject area, etc.

Also in 1964, a training program was conducted in the Catalog Branch for Library personnel.

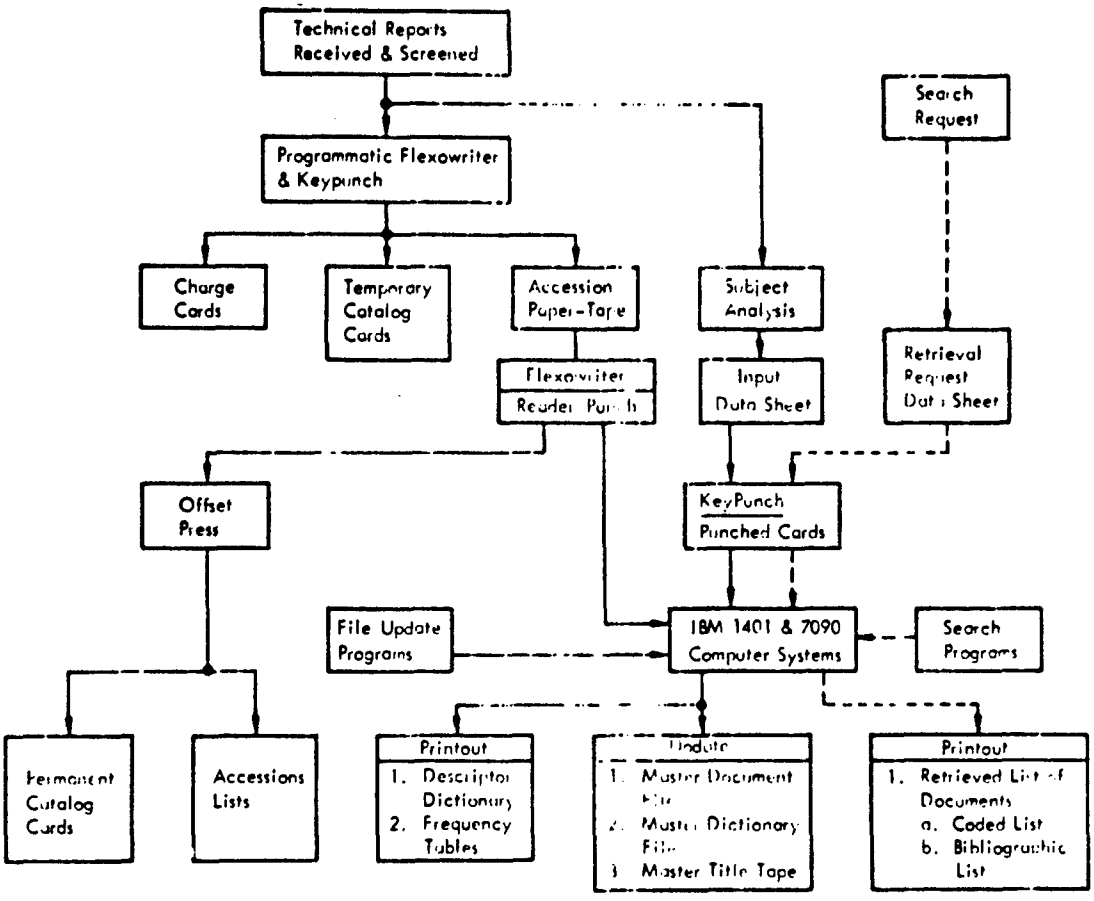
2. DESCRIPTION OF PROCESSES

Figure 1 depicts the information storage and retrieval system employed by the NOL Library for technical reports. The initial handling of technical reports (receipt and screening) is common to all functions (storage, retrieval, and dissemination and control) and proceeds as follows:

Step 1. Technical reports are received in the Catalog Branch daily. They are batched and prepared for processing daily. The reports are opened and the front covers are date stamped. If a mailing label has an "attention" line directed to an individual or area, it is cut

FIGURE 1

Flow Chart of Information Storage and Retrieval System for Technical Reports



off and stapled to the cover of the report. Reports with a classification of Secret are kept separate from the others. Receipts received with reports are forwarded to the branch supervisor for signature and returned to the Receiving Clerk for mailing to the respective report originators.

Step 2. The reports are then given a quick scan for extracting interlibrary loans or return of interlibrary loans and material before further processing of the batch. A check for the attention line is made on the inside of the report if it is not recorded on the mailing label. The classification group number is checked on each report. All Secret reports are logged. If a group number is omitted, the Catalog Branch forwards a form to the Security Division citing this violation by the originating activity. The Security Division takes the necessary action to resolve the security violation. If the subject material of the reports is of no special interest to NOL, the reports are discarded. If changes and insertions are received on reports, they are pulled for immediate updating of the report file by a Catalog Branch employee.

Step 3. This is a sorting function of the reports remaining in the batch. Reports are now sorted into permanent and temporary accession categories. At this point, a file is searched for documents

requested by staff members. If a requested document is received, the associated request form is pulled and attached to the document, which will be sent to the requester after processing. The reports rejected for permanent or temporary accessions are forwarded to interested parties or discarded.

Step 4. All permanent accessions are searched to avoid duplication of existing reports and assignment of a new accession number to subject material that is already on file.

Step 5. Subject headings such as Plastics, Physics, Engineering, Hydrodynamics, etc., are then assigned and are affixed to both the temporary and permanent accessions. The permanent and temporary reports are also assigned numbers for the NOL accessions list.

After this preliminary work is completed, the major areas of the NOL Library's information storage and retrieval program are carried out as follows:

(1) Information Storage

This function is comprised of the updating of the Master Document File, Master Dictionary File, and the Master Title Tape.

1. Updating the Master Document File and Master Dictionary File

The permanent reports undergo subject analysis in the Catalog Branch and are subject-coded on Form PRNC-NOL-5070/24-IBM 704 Work Sheet - Library Division. Reports which are of paramount interest to Laboratory personnel are retained as part of the permanent collection and are subject coded before being sent to the research personnel. Reports of lesser interest are first routed to individual technical personnel for evaluation. The technical staff members return their individual evaluations, including the addition of pertinent descriptors. The IBM 704 Work Sheet (input data sheet) now contains all the descriptors and codes relative to the report's subject matter. Appendix A shows the worksheet and provides instructions for its use, including definition and use of descriptors, codes, and code dictionary.

Next, the data on the IBM 704 Work Sheet are keypunched onto cards. Appendix B shows a sample of a completed worksheet, provides instructions for keypunching, and shows samples of the resultant punched cards. The forms and cards are then reviewed for transcription errors.

When this function has been completed, the worksheets are filed and the punched cards are forwarded to the Computer Applications Division (Mathematics Department) for updating of the Library files. The worksheets are destroyed after the Library files have been updated.

The punched cards are processed off-line onto magnetic tape via the IBM 1401 computer. The tape is then transferred to the IBM 7090 computer, and the Master Document File and Master Dictionary File are updated. New descriptors and codes that have been established are punched on cards and added to the Master Dictionary File at the time of updating of the Master Document File. On the punched card, dictionary codes are punched in columns 1-6, and the descriptor and identifying information are punched in columns 7-80.

In updating the Master Document File, the dictionary of descriptor codes is searched before the document data are added to the Master Document File. If a code term cannot be found in the Master Dictionary File, all information cards associated with all reports containing the illegal descriptor are rejected.

At the termination of the update run, a listing is produced which contains the numbers of those reports associated with the rejected descriptor. The listing is sent back to the Library for checking and later correction of the input data.

In addition to rejection of codes not listed on the dictionary tape, the program will reject input cards that do not conform to the update program format. Incorrect descriptor code count will also cause rejection. The deletion or withdrawal of reports from the Master Document File is also acoomplished by the same update program. For detailed information on the file update (maintenance) program, see Section III.

2. Updating the Master Title Tape

The Master Title File contains descriptive cataloging information on approximately 7,000 reports. (The Title File contains only those reports cataloged since July 1963.)

The input for this updating is obtained as a by-product of the descriptive cataloging performed on all permanent reports (see section on Information Dissemination and Control). It consists of a continuous punched paper tape

that contains bibliographic data on all new permanent reports processed for the week. The continuous tape was obtained during the processing of the daily permanent paper tapes for production of the accessions list. The paper tape is forwarded by the Library to the Computer Applications Division, which transfers the data to magnetic tape via the off-line 1401. A printout is then produced and sent back to the Library for checking and editing of the raw data. If a correction is required, a new entry is added to the magnetic tape and a delete card is prepared for removal of the undesirable data at the time of the update run.

Initially, delete cards were used to delete documents from the Title File which had duplicate numbers. Because the Library had not assigned unique numbers to documents until the past year, there existed in the collection a quantity with the same document number. The Library now assigns a unique number to all documents, and all entries in the Title File have a unique number. Prior to the update run on the 7090, a printout of the raw and edited data is usually obtained for final checking by the Library. During the update run, a listing is produced of all deletions made to the file. This listing is also used as a check on the Title File.

(2) Information Retrieval

This function deals with the requests for information and the ensuing machine search.

Reference requests made by NOL or contractor personnel which necessitate a subject approach require a machine search. A search request form (Figure 2) is made out by the reference librarian using the code terms (listed in the Code Dictionary) which in his opinion will best satisfy the reference request. Information on the search request form is punched on an IBM card.

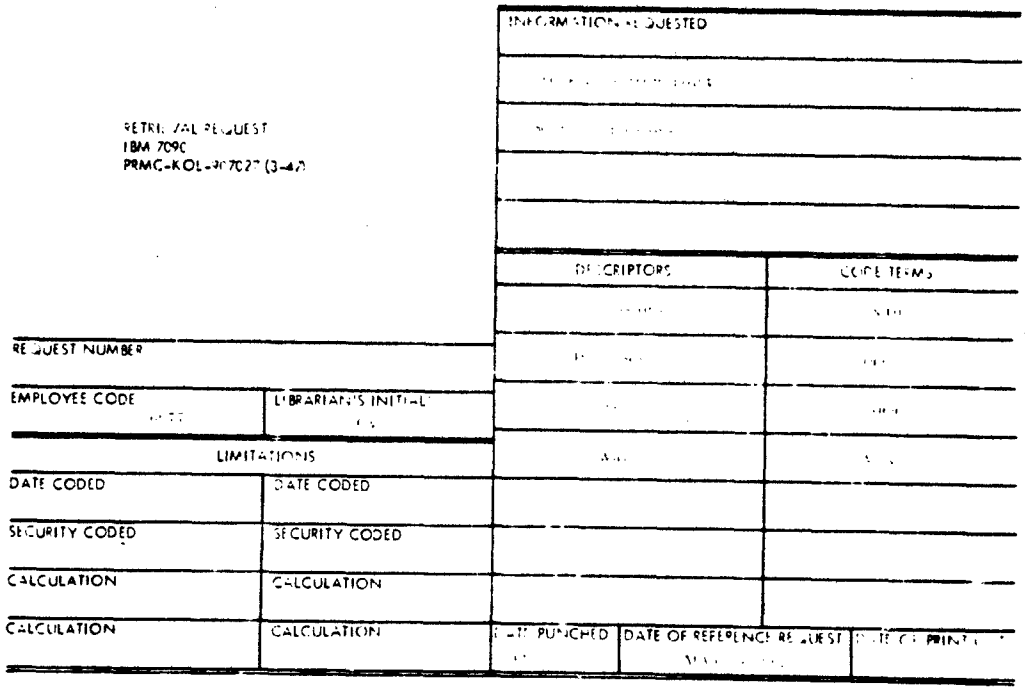


FIGURE 2
Search Request

All query cards for search requests are accumulated and forwarded to the Computer Applications Division for processing once a day. The Library has been guaranteed daily search time on the 7090.

The computer first searches the Master Dictionary File for code terms required by the search. If a descriptor code is not found in the Master Dictionary, the query is rejected. For those queries, all of whose descriptor codes are on the dictionary tape, the computer then searches the Master Document File for the information required. When the search does not find any reports that contain the required codes, the printout for this particular query number will read "No relevant documents found."

Individual searches are made on one to seven code terms. For a particular document to be retrieved during the search, all the code terms listed in the query must be included within the code terms contained by the document.

If information is desired on five code terms but there is a possibility that the Library has no material on all five terms, a separate search request form will be made out for any number

less than five that will provide acceptable information. Multiple searches may be made for a single reference request. The NOL Library Computer program permits 50 simultaneous searches.

Searches may also be made which permit the elimination of a particular descriptor code. Punching the code "NOT" in the descriptor code columns, followed by the descriptor code to be eliminated, will cause the computer to reject reports with the undesired code. Thus a search is made for reports on "Burning of solid propellant fuels but not (solid propellant) rocket motors" as follows:

Burning	BURN
Solid Propellant	SFUL
Not	NOT
Rocket Motors	ROCK

The Library computer program permits searches which are limited by sources, date of report, security classification, and circulation limitation. Thus, a request may specify limiting the search to relevant documents issued by a specific agency (such as NOL). Similarly a search request may specify only unclassified reports or reports classified no higher than Confidential. The computer search will eliminate unwanted reports on the basis

of security classification. Date limitations require the computer to eliminate all reports that do not conform to the specified dates. Thus the year "59" in the date limitation column will restrict the search request to reports issued later than 1959. By a similar pr ,edure, circulation limitation NOFO (not releasable to foreign nationals), REDA (Restricted data), and VTFU (VT fuze report) will not retrieve reports that have the specific circulation limitation associated with a report.

For reference requests that result in the listing of a large number of reports, the listing is furnished to the staff member and he selects the reports that he wished to consult. If the reference request resulted in only a very few reports being listed, the reference librarian will scan the reports for the information desired. For reference requests where complete source and/or title identification, series, or contract number is available, it is often faster and more economical to do a manual search in the card catalog. The conventional library card catalog is maintained for ready reference by source, title, personal author, series, and contract and project numbers.

A printout of report numbers associated with each search request is furnished the Library within two hours after completion of the searches. Two printouts are now produced. The first output, which consists of the input data from the query combined with the list of documents retrieved for that query, is shown in Appendix C-1. The second output, which is the result of running the list of retrieved documents against the Title File in order to obtain the title of the documents and other bibliographic data, is shown in Appendix C-2.

(3) Information Dissemination and Control

This function consists of the production of catalog cards, charge cards, the accessions list, descriptor dictionary, and frequency tables, and the control of storage via microfilming.

1. Production of Catalog Cards, Charge Cards, Accessions Lists

The descriptive cataloging of the permanent and temporary reports is prepared separately on a programmed Flexowriter which is coupled to an IBM 026 keypunch via a Friden Selectadata unit. The machine processing renders the following:

- six-part, card file forms
- two punched paper tapes (one permanent accession tape and one temporary accession tape)
- punched charge cards for each report (two cards for Unclassified and Confidential reports and three cards for Secret reports).

The six-part forms and the punched cards are reviewed line by line for transcription errors. The forms of both permanent and temporary categories are then sorted as follows:

Green Copy	By Issuing Agency
White Copy	By Document Number
Yellow Copy	By Document Title
Pink Copy	By Document Number
Blue Copy	By Series
Gold Copy	By Subject Category and then by Accession Number

When the forms have been completely sorted, all copies except the gold copy are taken to the Reference and Acquisition Branch and filed in the catalog file (see Appendix D for sample card).

The white copies of the form are stapled to the pink copies. The gold copy is retained to reflect any corrections when running the punched paper tapes for the accessions list.

The permanent and temporary reports with charge cards are prepared for final processing in the Catalog Branch. Charge card pockets are stapled to the back covers of reports and also to folders used to support reports that do not have substantial covers. The folders are completed by putting the reports inside for filing.

The punched charge cards are now inserted into the pockets of corresponding reports and folder-supported reports. When this function is completed, the circulation desk of the Reference and Acquisition Branch is called for pickup of the processed reports. Requested reports are routed to individual staff members, and the borrower's code--four digits of his payroll number--and the date of charge are entered on the card. (See Appendix E for a sample of charge card.) The remaining reports are filed.

The production of the accessions list is a weekly function and is not executed during the normal daily batch processing of reports. Once a week, the daily punched paper tapes of permanent and temporary accessions are taken from storage and are processed on a Flexowriter for preparing offset press paper masters.

During the processing of the daily permanent punched paper tapes, a continuous tape of descriptive cataloging for permanent reports is processed for the week to forward to the Computer Applications Division for updating of the title tape containing bibliographic information. The paper masters are processed on an offset press. Card stock is used to produce catalog cards for permanent reports to replace the temporary file copies in the card catalog. Paper stock is used for the accessions list of both permanent and temporary reports.

The catalog cards are taken to the Reference and Acquisition Branch and substituted for the file copies in the card catalog. The accessions list is distributed to staff members for review and analysis. Appendix F gives an example of the entries contained in an accessions list.

2. Descriptor Dictionary

A computer-produced listing of descriptors and codes, arranged alphabetically by code, is available in the Library for the use of Library patrons. This provides descriptor identification to enable the patron to interpret the computer printout that he receives in answer to his search request. The descriptors in the dictionary are restricted in their usage to areas of Laboratory interest.

There are now approximately 22, 000 descriptor codes in the dictionary. Of this number, 7, 500 descriptors are used for subjects. The remainder cover equipment, trade names, projects, geographic locations, corporate authors, personal names, etc.

Appendix G describes the development of descriptors and computer codes. Appendix H, descriptors and computer codes used in the NOL retrieval program, gives examples covering various categories.

3. Frequency Tables

Frequency tables for descriptor codes are maintained on the computer. These tables are updated at the time the weekly additions are made to the Library tapes.

A printout of frequencies, giving the descriptor codes and the number of times each is used in the Master Document File, is available at the reference desk. This listing is often consulted by the librarians and patrons when making out a search request to ascertain how much material will be retrieved on a search.

Thus, the code for "torpedoes" would retrieve approximately 1,900 references, while a search for "pulsatrons" would retrieve one reference.

Therefore, prior analysis of the contemplated search request defines the scope of the problem to the staff members and enables a more efficient search of the subject desired.

4. Microfilming

In order to conserve space and reduce handling problems, all permanent reports are eventually transcribed onto microfilm. This process is contractor performed. Through fiscal 1965, some 150,000 reports have been transcribed onto microfilm at a cost of approximately $140,000. Each microfilm card contains from 1 to 96 frames.

If a report is on microfilm, this is indicated on the shelf card. No additional cards or separate cataloging schemes are used for the microfilms. On demand, Library personnel will obtain the appropriate microfilm cards for the user who can read them at his leisure on the microfilm readers. If the user desires full-page copies, he may reproduce them himself on the microfilm reader/printer. The microfilm cards themselves are not allowed out of the reading room. The document library is open from 9 a.m. to 3 p.m., with "emergency" service available at other times.

3. MAJOR PROBLEMS

Major problems in the development and operation of the NOL Library information storage and retrieval program were limited by careful early planning and the following agreements which were reached before the program was begun.

- The programming would initially be done by the Mathematics Department.
- The Library would be allowed machine time each day for the reference searches.
- The Library budget would be increased sufficiently to cover charges for machine time and use of Mathematics Department personnel.

- There would be no increase in Library complement.

- No attempt would be made to analyze reports that were already in the Library and had been completely cataloged. (The exception would be NOL-originated publications.) After the program was in operation, the decision was made to analyze reports from 1958 on, in order to have all reports received between 1958 to 1963 analyzed and available in the computer.

- The present card catalog would be retained as an adjunct to information in the computer.

- The card catalog for source, title, personal author, series, and contract and project numbers would be maintained for all incoming reports. Subject information only would be available on the computer.

- Training of Library personnel to code reports would require time, and the transition from subject catalog to machine retrieval would be slow.

Problems have been reduced also for the following reasons:

- Personnel has not changed.

- Library staff did not pressure or push the mathematicians.

- Management backed the program.

- Programmers were given recognition for their work and have written their own reports on the Library computer programs.

- Reference and cataloging people rotate so that the same people see coding from both the input and searching aspects.

However, during the first five years of operation, a number of problems in the Library computer program have been encountered. Some of these problems were solved by changes in the Library procedures or in the computer program.

(1) It was found that a larger number of bound terms was needed than was originally anticipated in order to keep false drops to a minimum. Also required was a means of distinguishing object and agent. For example, in a search for aircraft submarine detection, is the aircraft detecting the submarine, or is the submarine detecting the aircraft?

To solve this problem, a fifth code letter was added to the original four-letter code, with A for Agent and O for Object. Thus:

AIRCA - Aircraft as Agent

AIRCO - Aircraft as Object

SUBMA - Submarine as Agent

SUBMO - Submarine as Object

(2) Some problems involve the usual difficulty of learning what the Library patron is looking for. This may be due to poor communication, reluctance on the part of the Library patron to

explain what he is interested in for security reasons, or lack of understanding of computer operation by the Library patron and the fact that specific information would assist the reference librarian in making the search more efficiently.

Often the Library patron insists that his search be broad--everything on "sonar detection," for example--where several hundred items would be listed on the printout. Or, he may insist that his search be too specific, thereby finding no reports answering his request.

The Library was recently asked to do a computer search on glass--no further restriction. The search was made, and the printout of several hundred items was given to the materials engineer. Later, it was learned that the engineer was working on a military classified project--since declassified--on the use of glass in place of steel for submarines.

(3) Another problem is providing additional report identification and abstracts of the reports retrieved on a search. A title program has been initiated which provides full bibliographic identification of report numbers on search printouts. However, this is limited to the 7,000 reports entered in the Title File since its inception. Input to the Title File is from punched paper tape

prepared on the Flexowriter at the time the descriptive cataloging is done. A procedural change in assignment of Library accession numbers for technical reports was required. At present, the Library Document File uses accession numbers to identify each document. Parts, volumes, and supplements, therefore, do not have a unique number. The Title File, however, requires that each document have a unique number.

Two methods for assigning unique numbers to technical reports were studied. The proposed methods were as follows:

- All technical reports would be assigned a unique number, and this would replace the present use of accession numbers in all Library operation; or
- Unique numbers would be assigned for computer use only, with printed tables of equivalents where the same Library accession number is used for parts, volumes, supplements, etc.

The first method was chosen and implemented.

4. ACTIVITIES BEING PLANNED OR DEVELOPED FOR MECHANIZATION

NOL investigated the possibility of providing abstracts with computer searches. This idea was rejected primarily due to the major increase in cost required for its operation.

Plans are under consideration for the preparation of a book catalog and an annual periodical listing from the punched cards. At present, 90 percent of all periodicals have been entered on punched cards, and it is expected to be completed during fiscal year 1966. NOL is considering cooperating with the Applied Physics Laboratory of Johns Hopkins to produce an annual listing of periodicals available in technical libraries in the Washington-Maryland area.

III. PROGRAM SYSTEM DATA

The programs for the Library information storage and retrieval system fall into the following two major categories.

- Master Tape Maintenance - which is subdivided into a program to update the Master Document File and Master Dictionary File, and a program to update the Master Title File. (These files are described in Appendices J, N, and O.)
- Reference Search - which is a program that searches the Master Document File for reference documents requested by the technical staff.

1. MASTER TAPE MAINTENANCE

(1) Document File and Dictionary File Update Program

The file maintenance run for the Master Document File and the Master Dictionary File is depicted in the system block diagram of Figure 3.

The inputs to the program--new descriptors, new documents, and document deletions and I. D. changes, all described in Section II--are placed on magnetic tape via the 1401 (off-line).

FIGURE 3

Update of Master Dictionary and Master Document Files

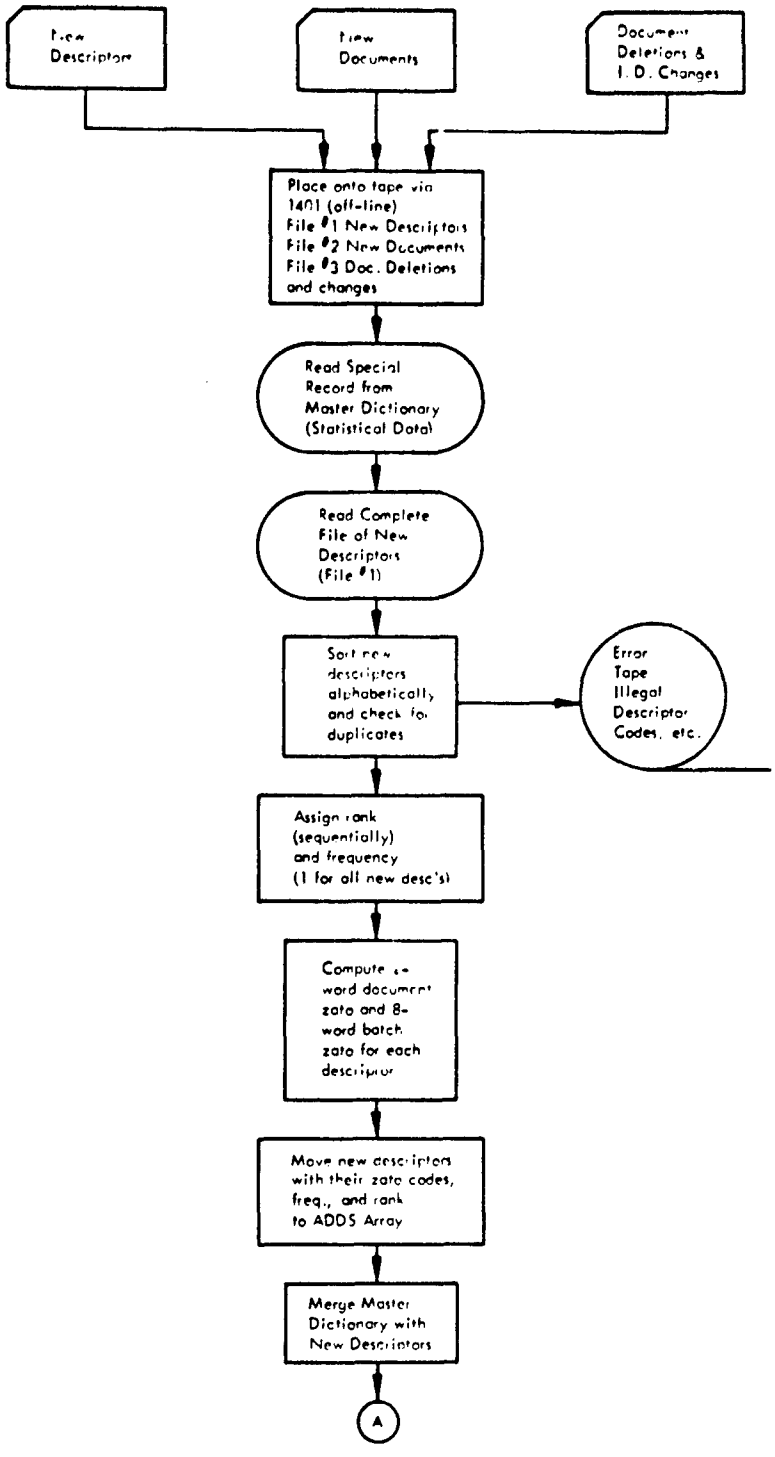


FIGURE 3 (cont.)
Update of Master Dictionary and
Master Document Files

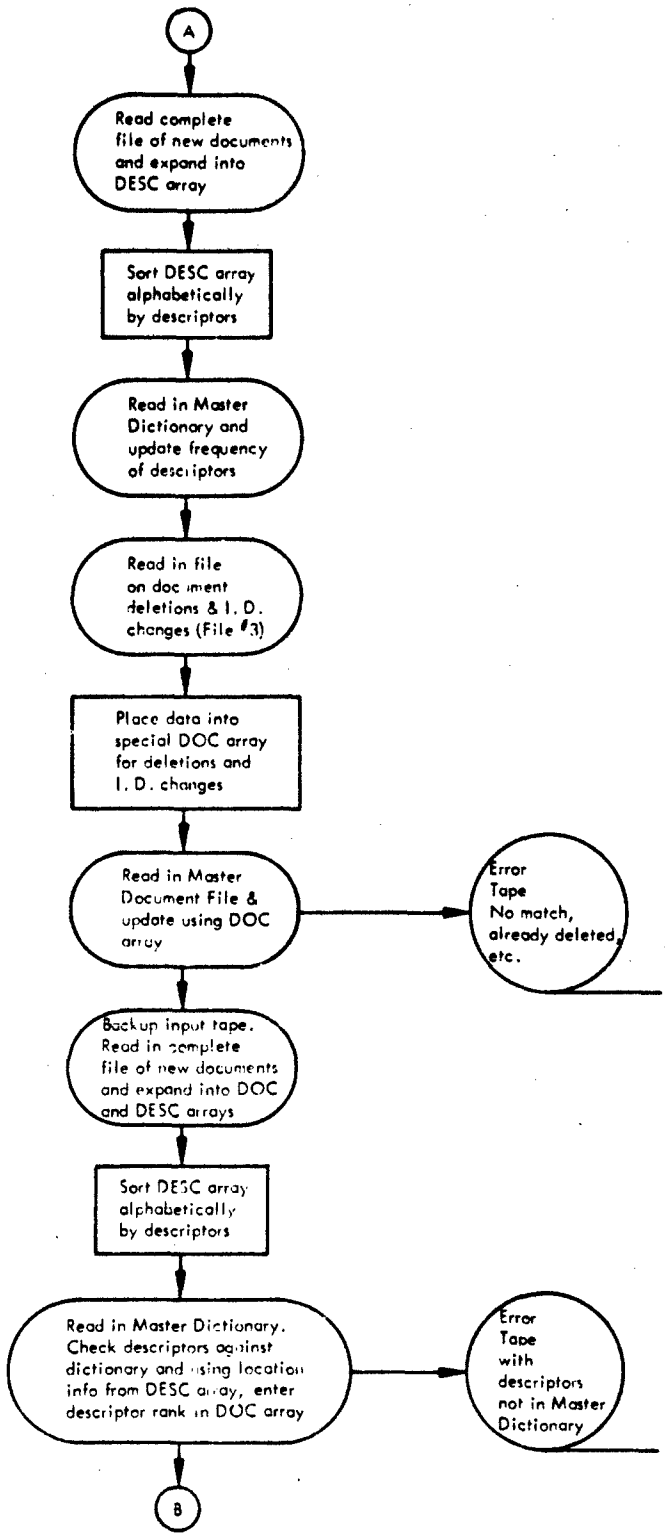
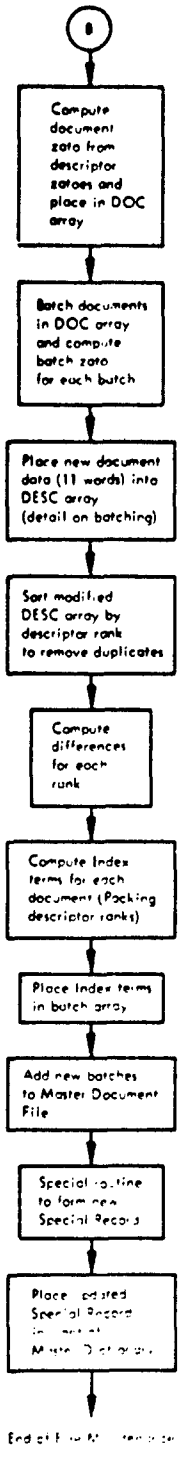


FIGURE 3 (cont.)

Update of Master Dictionary and Master Document Files



The input tape is made up of three files. The first file consists of all new descriptors. The second file consists of all new documents. The third file contains all document deletions and I. D. changes.

The program starts by reading the first record (Special Record) of the Master Dictionary File (see Appendix I for format of Special Record, and Appendix J for format of Master Dictionary Record), and then proceeds to read the complete file of new descriptors from the input tape. The descriptors are then sorted alphabetically, duplicates are rejected, and ranks are assigned sequentially beginning with the number following the last assigned rank in the dictionary (number obtained from Special Record). A number indicating frequency for each descriptor is also assigned and is assumed to be one for all new descriptors.

The next step is to compute for each new descriptor its eight-word (batch) and two-word (document) "zato" code. This code, which is a pattern of random digits, is later used to represent the term durirg the primary searching operations (see Appendix K for further explanation). As the codes are computed for each descriptor, the descriptor and its associated zato codes, rank, and frequency are formed into an array in core called

the ADDS array (format similar to dictionary record, see Appendix J). The Master Dictionary is now merged with the new descriptors by use of the ADDS array.

Next, the complete file of new documents is read in from the input tape. The descriptors are stripped from the document entries and placed into the DESC array. The DESC array is then sorted alphabetically by descriptor. The Master Dictionary is then read in, and the descriptor frequency in the Dictionary is updated.

The next step is to purge the Master Document File prior to adding the new documents. This is accomplished by making deletions and I. D. changes. The last file on the input tape (file #3) is read in, and the data are placed into a special DOC array for deletions and I. D. changes (see Appendix L). The Master Document File is then read in and updated using the special DOC array. Illogical items such as deletions of documents not in the Master Document File are rejected and output onto an error tape.

After the Master Document File has been purged, the input tape is backed up to the new documents file. The complete file of new documents is then read in from the input tape and

expanded into a DOC array (see Appendix L) and DESC array (DESC array consists of two words for each descriptor: first word-descriptor, second word-location of document from which descriptor was obtained). The DESC array is then sorted by descriptor. The Master Dictionary is then read in, and the descriptor rank for each descriptor in the DESC array is placed in the DOC array. Descriptors in the DESC array for which a matching descriptor in the dictionary cannot be found are rejected. The descriptor and the associated document are deleted and output onto an error tape. Next, for each document in the DOC array, the document zato is computed by superposition of each of the zatoes of its associated descriptors. After the document zatoes are computed and placed in the DOC array, the documents are batched (normally six documents per batch) and the batch zato codes computed. The batch zato is computed by superposition of each of the zatoes of its included documents.

The program continues by placing into the DESC array, for each new document, the document's two-word zato and nine words of indicative information. The modified DESC array is then sorted by descriptor rank, and all duplicates are removed. Then to conserve storage, the ranks are differenced, and the differences are packed into index term numbers (see Appendix M).

The index term numbers for each document are then placed at the end of the batch in which the document is contained. The new batches are then added serially to the Master Document File (see Appendix N for format of Document Record). At the end of the update, a special routine is called in to update the Special Record.

(2) Title File Update Program

The file maintenance run for the Master Title File is depicted in the system block diagram of Figure 4.

The input to the program is a continuous paper tape of permanent new reports processed for the week. The continuous tape was obtained during the processing of the daily permanent paper tapes for production of the accessions list. The tape contains the exact data produced during the descriptive cataloging of the technical reports.

The data are transferred to magnetic tape via the off-line NOL Data Conversion System (this tape is referred to as the Stacked Title Tape--see Appendix O for format). A printout of the raw data is obtained from a 7090 run for checking and editing. If a correction is required, a new entry is punched on tape and added to the Stacked Title Tape, and a deletion card is prepared

FIGURE 4
Update of Master Title File

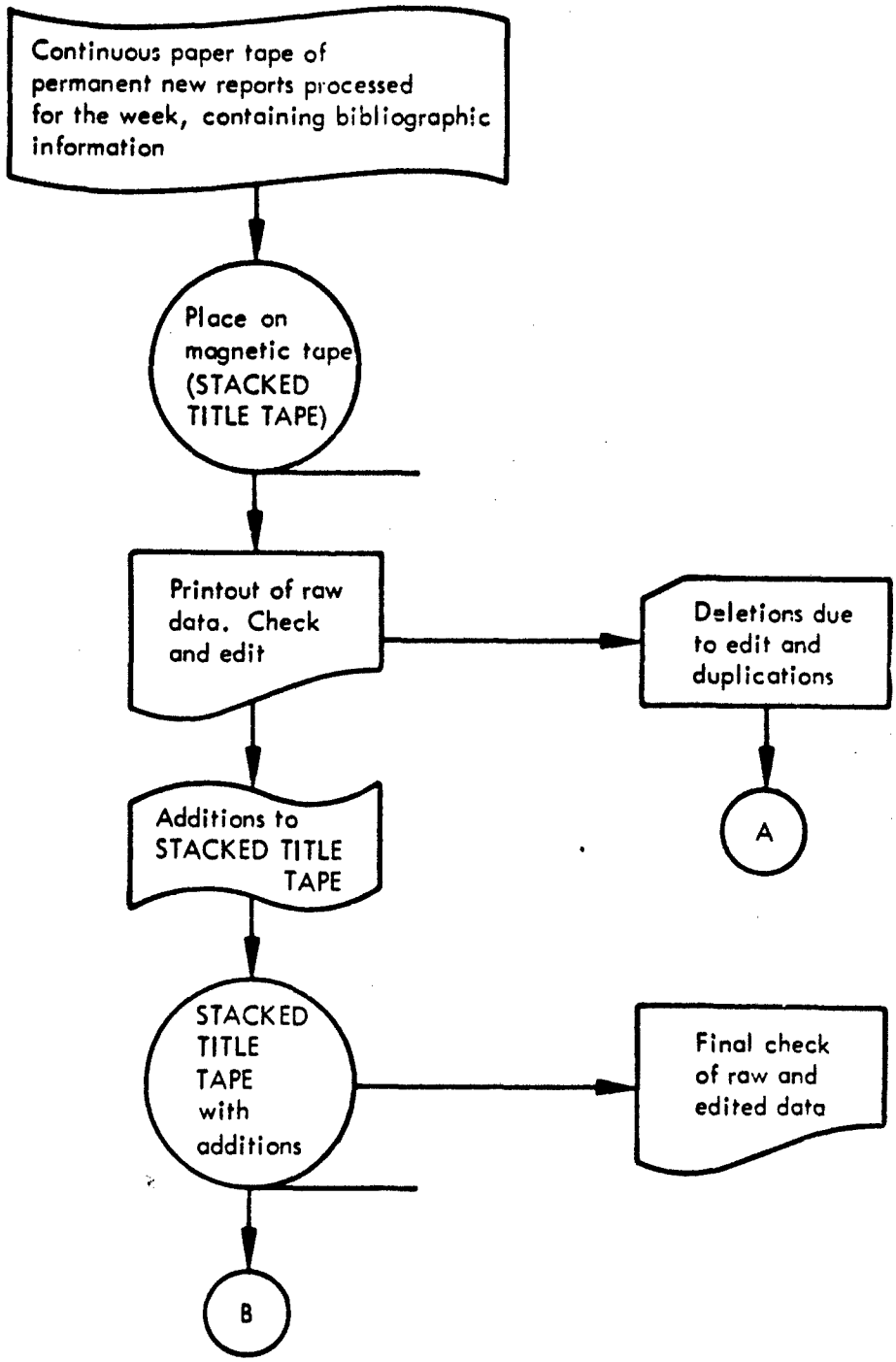
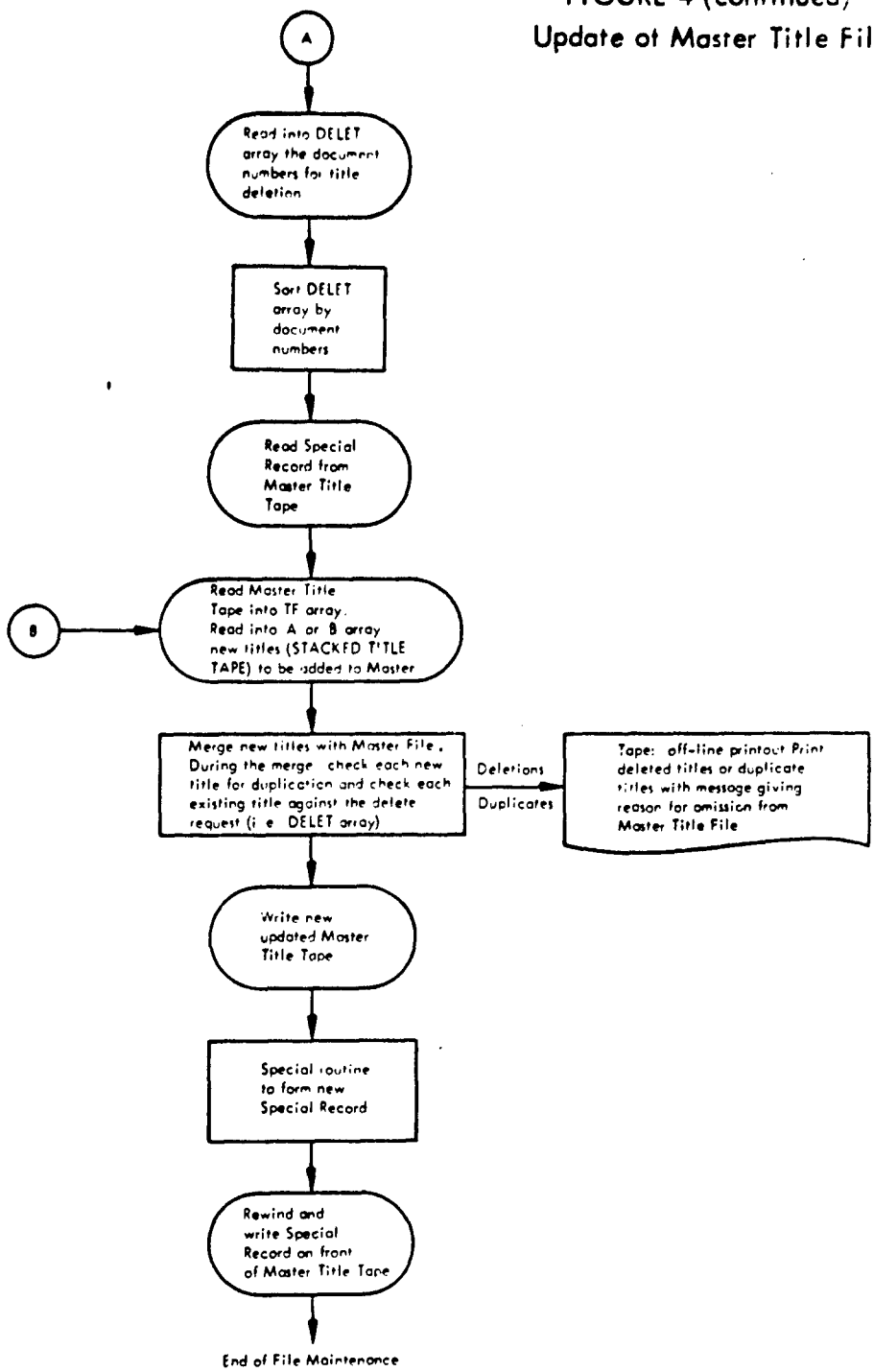


FIGURE 4 (continued)
Update of Master Title File



to remove the undesired data. Deletion cards are also used to delete documents with duplicate numbers from the Title File. This situation exists because the Library had not assigned unique numbers to documents until the past year. Therefore, there may be any number of documents with the same number. Any documents with the same number will eventually be given a unique number or will be excluded from the Master Title File. Prior to the update run on the 7090, a printout of the raw and edited data is usually obtained as a final check.

The program begins by reading into the DELET array (see Appendix O) the document numbers for title deletions. The DELET array is sorted by document number. The Special Record from the Master Title Tape is read in, and then the Master Title Tape is read into core as a TF array (same as Title File Record, see Appendix O). The Stacked Title Tape is then read in and formed into an A or B array (see Appendix O).

The new titles are now merged with the Master File (TF array merged with A or B array). During the merge, each new title is checked for duplication, and each existing title is checked against the delete requests (DELET array). Duplicates and

deletions, together with a message giving the reason for omission from the Master Title File, are output onto a separate tape for later printout off-line. The result of the merge is an updated Master Title File. A special routine is then called in to update the Special Record. The Master Title Tape is rewound, and the Special Record is entered on the front part of tape.

2. SEARCH PROGRAM

The search run is depicted in the system block diagram of Figure 5. The input to the program (query cards) is placed on magnetic tape via the off-line 1401. The program starts by reading in the query file and forming the query array (same format as card). The descriptors from all queries are then pooled together into the DESC array (DESC array consists of two words for each descriptor: first word--descriptor, second word--location of query from which the descriptor was obtained and sign to show if the descriptor is negated). The DESC array is then sorted by descriptor.

Next, the Special Record is read in from the Master Dictionary, and the security, date, and circulation limits are coded (using tne tables contained in the Special Record) and placed in the query array. The Master Dictionary is now read in, and all of the descriptors in the DESC array are compared against those of the dictionary. When a

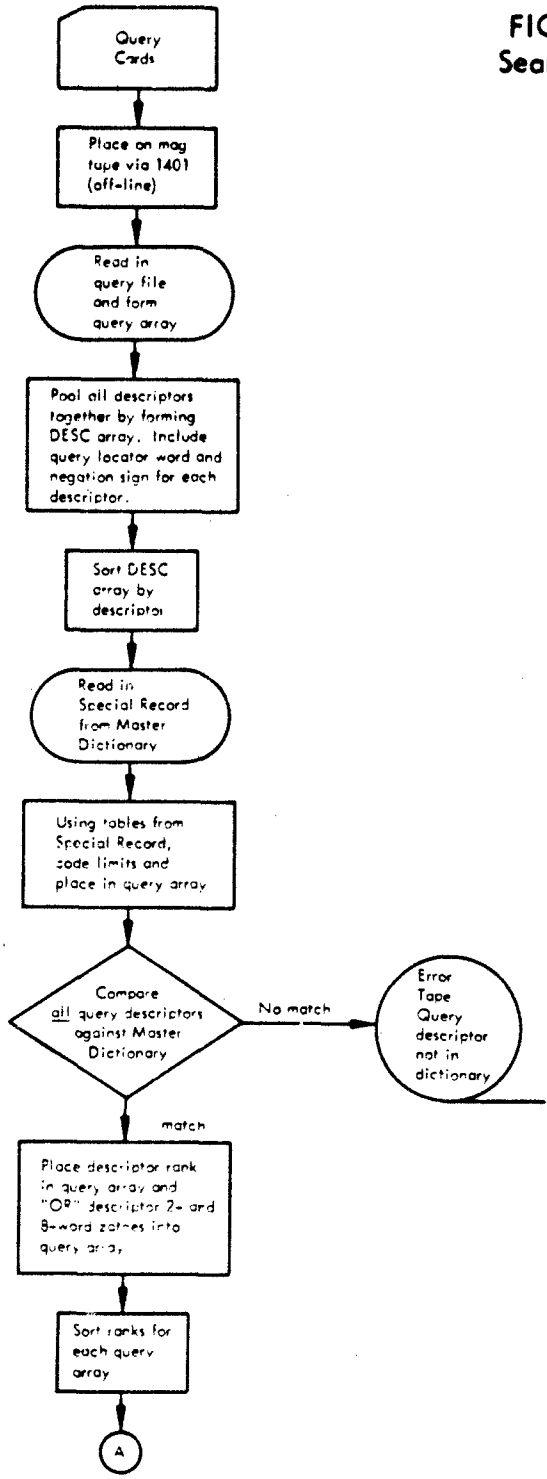


FIGURE 5
Search Run

FIGURE 5 (cont.)

Search Run

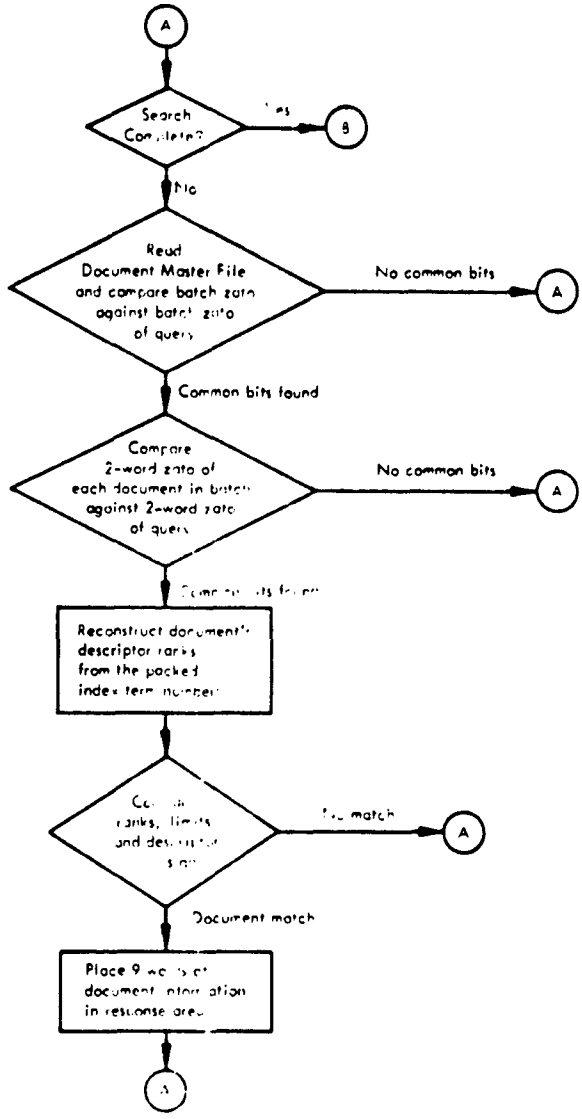
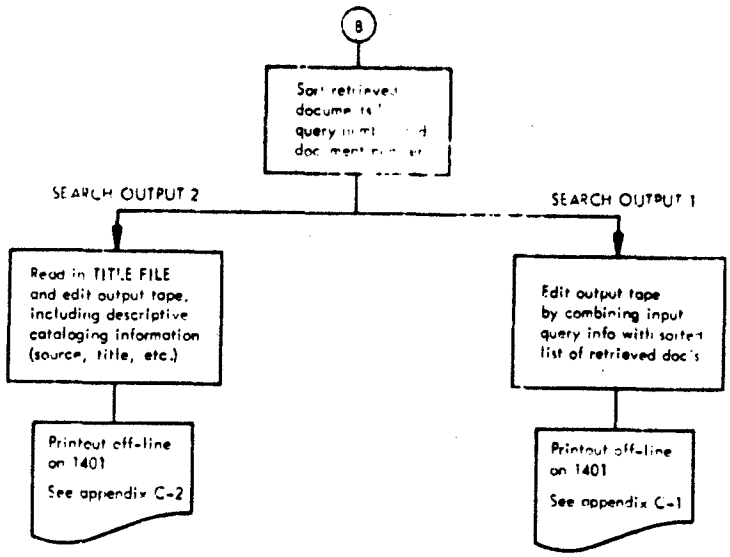


FIGURE 5 (continued)
Search Run



match is found, the dictionary descriptor's rank is placed in the query array under the query (or queries) from which the descriptor was obtained. Also, the descriptor's eight- and two-word zatoes are superposed into that descriptor's query zato area. Any query with descriptors not appearing in the dictionary or errors in its limits is rejected and output onto an error tape. At the end of the pass, the ranks within each query are sorted.

The actual search now begins. The Master Document File is read in, and the batch (eight-word) zato in each batch is compared against the batch zato from each query. If no bits are found in common between the codes, the Master Document File batch is passed without further inquiry, and the next batch from the file is read in. If common bits are found, the program drops to the next level of scan. The two-word document zato from the query, which has common bits with the batch, is compared with the two-word zato of each document in the batch. Again, if no bits are found in common, the program proceeds to the next batch. If common bits are found between the two-word zatoes, the program drops to the next level of scan.

The next level consists of first reconstructing the matched document's descriptor ranks from the packed index term numbers, and then comparing ranks, limits, and descriptor sign. For a document

to be accepted in answer to a query, it must contain all positive descriptors from that query and no negated descriptors. It must contain the specified circulation limits and lie on or within the data and security limits. If a document match is obtained, the nine words of document information are then placed in the response area. The search continues until up to 2,000 documents are retrieved on any one query, or a maximum of 2,111 for the total search.

After the search is completed, the retrieved documents are sorted by document number and then by query number. Two outputs are now obtained. The first output, which consists of the input data from the query combined with the list of retrieved documents for that query, is shown in Appendix C-1. The second output is the result of running the list of retrieved documents against the Title File and obtaining the title description and other bibliographic data for each retrieved document which is also contained in the Title File. This output is shown in Appendix C-2.

IV. EQUIPMENT AND COSTS

For the information storage and retrieval system, the Library rents machine time and programming services from the Mathematics Department. For fiscal 1965, these services were valued at approximately $12,000. These charges will continue to decline since all major programs have been completed and changes or generation of other programs is now being handled by the Library's own programmer/librarian.

Unit charges to the Library for machine time are $60 an hour for the IBM 7090 system and $10 an hour for the IBM 1401 system, which is mainly used as an off-line peripheral processor for the 7090. Average weekly usage is 4 hours for the 7090 system and 2 hours for the 1401 system. In addition, another 50 hours per year of 1401 system time is approximated as required for the printing of the descriptor listing which is updated quarterly.

The 4 hours indicated for the 7090 includes processing of input information for new reports as well as time for reference searches. At the present time, the average cost for a single search (covering computer time only) is estimated at $2.00. The $2.00 is due to the

addition of the Title Program, which increases machine time and is dependent upon the number of titles in the Title File for each search. Prior to the addition of the Title Program, the average cost per search was estimated as $0.80.

The Library has assigned 11 full-time people to functions associated with Library automation. The grades for the 11 people, including one programmer/librarian, vary from GS-6 to GS-12.

The Library owns and operates the following equipment:

(1) Two Friden Programmatic Flexowriters with tab card punch control and Selectadata units. The original cost of each machine was approximately $6,000. Total machine usage averages 14 hours per day. There are three people, grades GS-4 to GS-6, trained to operate this equipment.

(2) Two IBM 026 Keypunches costing $3,400 each. Their total average daily use is 16 hours. There are seven people, grades GS-4 to GS-6 who are capable of operating this equipment.

(3) Four Remington Rand Kard-a-Film Readers. These readers, costing approximately $150 each, are used to read the reports on microfilm.

(4) One 3M Thermofax reader/printer, model Filmac 100. This reader/printer, costing approximately $700, is used for reading microfilm or obtaining regular-size page reprints of the reports on microfilm. The Library has on order an additional reader/printer, model Filmac 400 which will cost approximately $1,200.

The microfilm readers and printers are used and operated directly by the NOL staff; therefore, usage rates are not available. The cost of microfilming the 150,000 reports, available through fiscal 1965, was approximately $140,000.

Other costs attributable to automation include the printing of the weekly accessions list for documents. The listing, which is reproduced from offset press paper masters generated on the Flexowriters, costs approximately $200 a month, including materials for the 125 copies produced. The Library effort involved is approximately 2 hours a day by one operator of a GS-4 level.

BIBLIOGRAPHY

Batch Processing of Technical Reports, Study by George F. Herbold, Jr.

Descriptors and Computer Codes Used in Naval Ordnance Laboratory Library Retrieval Program, NOLTR 64-20, AD 608404, 15 October 1964, U.S. Naval Ordance Laboratory, White Oak, Maryland.

NOL Retrieval System for the 7090, NOLTR 64-125, AD 609970, 1 October 1964, U.S. Naval Ordnance Laboratory, White Oak, Maryland.

Pioneering in Machine Literature Searching and Retrieval (U), NAVWEPS Report 7388, AD 255291, April 1961, U.S. Naval Ordnance Laboratory, White Oak, Maryland.

Procedures for Analyzing Technical Reports for Machine Retrieval, NOLTR 62-51, AD 278454, 9 May 1962, U.S. Naval Ordnance Laboratory, White Oak, Maryland.

APPENDIX A

INSTRUCTIONS FOR USE OF INPUT DATA WORKSHEET INCLUDING USE OF DESCRIPTOR CODES

(Extracted From NOLTR 62-51)

Changes since the date of publication of NOLTR 62-51 are not reflected.

APPENDIX A

INSTRUCTIONS FOR USE OF INPUT DATA WORKSHEET INCLUDING USE OF DESCRIPTOR CODES

The work sheet (figure 1) is used for listing identification information and subject analysis of the technical report. Identification information is inserted in the seven blocks across the top of the work sheet. Subject information utilizing descriptors and their codes are listed in two vertical columns below the identification information. A maximum of 24 descriptors is listed on each work sheet. If more than 24 descriptors are required, an additional work sheet is used. A minimum of two descriptors may be assigned to a technical report; the maximum number is without limit depending upon the subject content of the report.

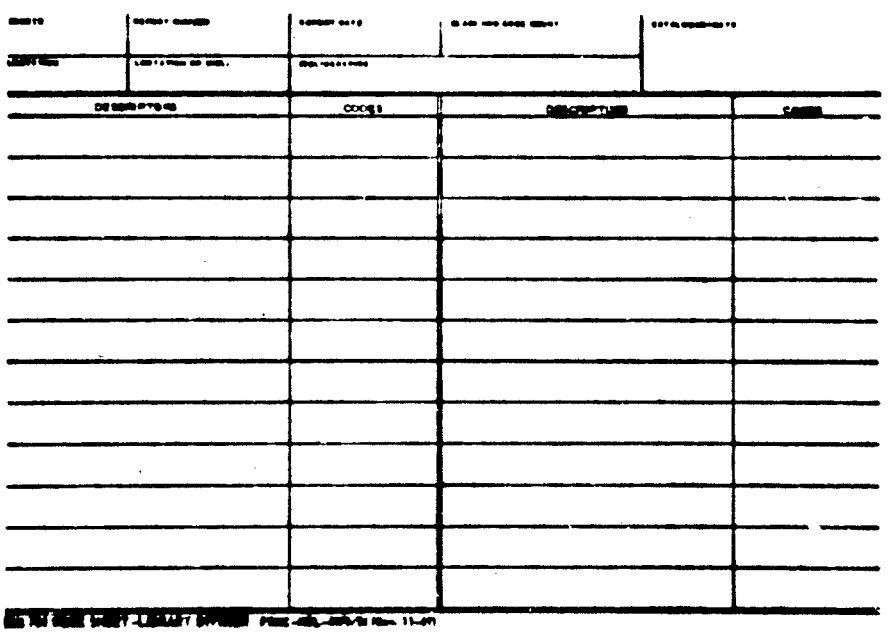


Figure 1 IBM Work Sheet

While two descriptors assigned to a technical report is rare, many reports pertaining to explosives, plastics, aircraft models and designations, may have 50, 75 or 100 descriptors. Work sheets may be handwritten or typed, so long as code terms are legible.

SOURCE IDENTIFICATION

The first block on the work sheet is used for source identification. This is the issuing agency responsible for the report -- a government agency or establishment, committee, university, industrial firm or an individual (not identified with a corporate body). When the source has been identified the Code Dictionary is consulted and the four letter code for source is filled in. If a code for a source has not been established, the Chief Cataloger establishes the necessary code.

For source identification the name of the Laboratory is used in those instances where the Laboratory is widely known without reference to the parent organization:

Ballistics Research Laboratory (Aberdeen Proving Ground) BRLA
Chesapeake Bay Annex (Naval Research Laboratory) CBAX
Defense Research Laboratory (University of Texas) UTDR
Ordnance Research Laboratory (Pennsylvania State University) ORDA
Jet Propulsion Laboratory (California Institute of Technology) JPLA
Applied Physics Laboratory (Johns Hopkins University) APPL
Lawrence Radiation Laboratory (University of California) UCRL

A subsidiary of a company is coded as an independent source:

Formica Co. (subsidiary American Cyanamid Co.) FRMC
Phillips Chemical Co. (subsidiary Phillips Petroleum Co.) PHCH
Aerojet-General Co. (subsidiary General Tire and Rubber Co.) AERJ

A division of a company is coded as an independent source if it has existed previously as an independent company:

Brush Instruments (Division Clevite Corp.) BRUH
Convair (Division General Dynamics Corp.) COVA
Kearfott (Division General Precision, Inc.) KEAR
Sperry Gyroscope Co. (Division Sperry Rand) SPRR

If the division has not existed as an independent entity, the parent organization is used as the source:

Nortronics code Northrop Corp. NPTN
Shell Development Co. code Shell Oil Co. SHLO
Sylvania Electronic Systems code Sylvania Electric Products, Inc. SYLV

For reports issued by the Naval Ordnance Laboratory, series designation of the report is retained as the source:

NOLTR	Naval Ordnance Laboratory Technical Report
NAVD	NAVORD or NAVWEPS Report (NOL-originated publications only)
NAVDOD	NAVORD OD or NAVWEPS OD (NOL-originated only)
NOLTN	Naval Ordnance Laboratory Technical Note

LIBRARY REPORT NUMBER

In the second block of the work sheet the report number is indicated. This is an accession number assigned by the Naval Ordnance Laboratory Library to each incoming report. In accordance with the computer program the report number is always expressed by six digits -- if the number of the report is less than six digits, seros precede the number to make a total of six. A diagonal line through the sero is used to distinguish the sero from the letter O:

91287	Ø91287
115718	115718
3257	ØØ3257

For NOL-originated publications, the series number is used as the report number:

61-101	61Ø1Ø1
7254	ØØ7254
61-1	61ØØØ1

DATE OF REPORTS

The report date is the publication date of the report. This is expressed by four numbers, using the month and year. Where the month of publication is unknown, the year only is used as date of the report. Months are designated by numbers 1 through 12 followed by two digits for the year. Months requiring a single number are preceded by sero in order to utilise four characters for the code:

June 1961	Ø661
Dec. 1961	1261
1960	ØØ6Ø

SECURITY CLASSIFICATION

Security classification plus descriptor count utilise four characters. Security classification is indicated by a single letter:

S Secret
C Confidential
C Restricted (foreign only)
U Unclassified
D Discreet (British Secret or Confidential)

The remaining three characters indicate the total number of descriptors assigned to the report. Here again four numbers are mandatory. For example, a secret report having 29 descriptors assigned will be SØ29; a confidential report having a total of five descriptors will be CØØ5; a British secret discreet report having a total of 40 descriptors will have the number DØ4Ø.

CIRCULATION LIMITATION AND BIBLIOGRAPHIC INFORMATION

The remaining three blocks are utilized for circulation limitation or bibliographic information. A report which has the category restricted data is indicated by REDA; formerly restricted data FRDA; not releasable to foreign nationals NOFO.

Circulation limitations are listed in Appendix A. A report may require one, two or three different circulation limitations.

Bibliographic information covers volume, part, supplement, appendix, etc. This is indicated in the last block if there is one bibliographic designation. If two bibliographic codes are necessary, the last two blocks are used. If three bibliographic codes are required, three blocks are used for bibliographic information. The total number of characters available for circulation limitations and bibliographic information may not exceed 14 characters. If the total number of characters required is more than 14, a decision must be made as to the relative importance of circulation limitation versus bibliographic information.

A list of descriptors and codes for bibliographic information is given in Appendix B.

DESCRIPTORS AND CODES

Upon completion of the identification information, the person analyzing the reports lists the descriptors and codes covering the subject content of the report. Subject analysis of technical reports requires analysis in depth, and both general and specific subject areas are covered. Descriptors assigned to a report vary in number, the average is 12-16 descriptors.

The following rules serve as a guide to the coder in analyzing the report:

1. Subject covered - what the report is about
2. Method covered - tests, design, quality evaluation, procedures
3. Factors discussed - properties, characteristics, effects, evaluation
4. Equipment used
5. Environmental factors - atmosphere, laboratory, climatic conditions, geographic location

The first four descriptors assigned to a report will be those terms associated with the title of the report. Where the title lacks subject content, pertinent terms which cover the subject area of the report are listed. Thus a report, "Hydrodynamic Characteristics of Subroc Missile" will have the first four descriptors listed in order of importance:

Subroc
Missile
Hydrodynamics
Characteristics

Descriptors which follow the first four descriptors are listed at random.

USE OF DESCRIPTORS AND CODE TERMS

Descriptors selected to cover the subject content of the report are defined to clarify their meaning and restrict their usage:

Currents (electric)	CURR
Currents (ocean)	CURE
Detector (electric or mechanical)	DETC
Detector (chemical)	DTEC
Waves (ocean)	WAVE
Sound waves	ACOU
Ballistics	BALS
Ballistic-type	BALI

A descriptor and its synonym use the same code:

Calculation	COMA
Computation	COMA
Airplanes	AIRC
Aircraft	AIRC
Apparatus	EQUI
Equipment	EQUI

Phrases which convey a single concept are treated as one descriptor:

Dead period	DEAD
Ground zero	GROZ
Angle of attack	ANGE
High frequency	HIGF
Jet stream	JETT
Kill probability	KILL

EQUIPMENT DESIGNATIONS

Since a large portion of the Laboratory program is concerned with ordnance hardware, equipment discussed in a report is always identified by name and model number. For equipment used by the Navy, Mod (modification) number is considered an integral part of the identification.

Mine Mark 51 is coded as follows:

Mines	MINE
Mark	MARK
51 Mod 0	51MØ

(If modification number is not indicated, it is assumed to be modification 0)

Due to the Navy use of Mark and Mod numbers for its equipment, it is necessary to differentiate between major pieces of ordnance and minor components. Failure to make this distinction in assigning codes will result in erroneous information on retrieval since a Battery Mark 51 may be used with the Mine Mark 25, Mine Mark 36, or Mine Mark 52. In coding a report on the Mine Mark 36 which has the Battery Mark 51 as a component, the descriptor codes selected would be:

MINE
MARK
36MØ
BATT
51MØN

When coding minor items, the letter N follows the modification number. Without the letter N for the battery number designations, a search for reports in the Mine Mark 51 would erroneously retrieve this report since the battery is designated as Mark 51.

Major ordnance items include mines, torpedoes, missiles, bombs, fire control equipment, etc. Components include batteries, fuses, detonators, primers, etc. Appendix C lists major and minor pieces of ordnance.

For equipment which does not employ the Mark number, there is no distinction made between major ordnance and minor components.

In coding British equipment which is identified by Mark and modification number, the modification number is disregarded, thus avoiding false drops, in searches for American developed equipment.

Testing equipment is identified by the Mark number disregarding the modification number. This eliminates erroneous retrieval or "false drops" where a test set is identified by Mark number and the ordnance with which it is associated has a different Mark and modification number. For example, a rocket is designated as Mark 22. The test set used to check its electrical system is Mark 37. Unless a distinction between the two pieces of hardware is made, a search for the Mark 37 rocket would retrieve the Mark 37 testing equipment as a "false drop":

Rockets	MISL
Mark	MARK
22 Mod 0	22MØ
Testing equipment	TESR
37	ØØ37 not 37MØ

COMPANY NAMES

A descriptor for the company name is used where the name is associated with the equipment.

Brush AX11	Brush AXAA ØØ11
Polaroid Land Camera	Polaroid Land Camera

USE OF RESEARCH, STUDY, DEVELOPMENT

The descriptors "Research", "Study", "Development" are used where they are associated with the title of the report or the report is general in nature and a small number of descriptors is required.

DESCRIPTORS AS NOUN, VERB OR ADJECTIVE

A single descriptor serves for noun, verb, or adjective form of a word with a separate descriptor for the instrument or equipment. Thus:

Detonation (noun)	DETO
Detonate (verb)	DETO
Detonating (Fuze) (adjective)	DETO
Detonators (instrument)	DETN
Acceleration (noun)	ACCL
Accelerate (verb)	ACCL
Accelerating (Voltage) (adjective)	ACCL
Accelerators (instrument)	ACCR

RULES GOVERNING NUMBER CODES

Due to limitations of the IBM computer operation, rules for indicating numbers and equipment designations were set up.

Four to six characters are required for each code. There may not be less than four or more than six characters. Where there are less than four characters, zeros are inserted to make a total of four:

1	ØØØ1
10	ØØ1Ø
100	Ø1ØØ
1000	1ØØØ
10000	1ØØØØ
100000	1ØØØØØ

Mark and modification numbers are indicated as follows:

1 mod 0	Ø1MØ
1 mod 10	1M1Ø
20 mod 12	2ØM12

Where a single letter precedes the number, code is indicated as follows:

A1	AØØ1
C10	CØ1Ø
B100	B1ØØ
K1000	K1ØØØ

Where two letters or three letters precede a number, separate descriptors are required. In each instance, the four letter descriptor code is arbitrarily selected from available alphabetic combinations.

Two letters followed by a number:

BA10	BAZZ
	ØØ1Ø
CD14	CDAA
	ØØ14
AT27	ATAA
	ØØ27

Three letters followed by a number:

BAZ1	BAZA ØØØ1
CDE10	CDEX ØØ1Ø
GEF100	GEFZ Ø1ØØ

Where a single letter follows a number, the letter is omitted. The letter designation usually indicates a slight variation in the equipment, and searches using the number only will retrieve all minor variations.

1A	ØØØ1
10B	ØØ1Ø
100C	Ø1ØØ
1000F	1ØØØ

Where a number precedes two or more letters, this is considered part of the designation and is retained as a necessary identification:

2BA	Ø2BA
10CDE	1ØCDE
100EF	1ØØEF

When a letter occurs between numbers, it is retained as part of the number identification:

A3F1	A3F1
3FA1	3FA1
3FD10	3FD1Ø

For aircraft designations the letter following a number is retained:

FJ1H	FJAA ØØ1H
AD3A	ADAA ØØ3A
B52H	B52H
FH102A	FHAA 1Ø2A

The letter X is used in the fourth position in order to use the same letter combination for the same series of aircraft designations:

A3D0	A3DX
A3D1	A3D1

Numerical designations for equipment which include a hyphen (-), period (.), or diagonal (/) replace the hyphen, period or diagonal with the letter X where it is an integral part of the number:

20-20	20X20
3.54	3X54
A17/2	A17X2

The hyphen or diagonal is ignored when it is used merely to separate a letter from a number:

A-10	AØ1Ø
10-EF	1ØEF
B/21	BØ21

Mach numbers are indicated as follows:

Below Mach 1:

.39	ØX39
.63	ØX63

Above Mach 1:

1 - 1.24	1XØØ
1.25 - 1.49	1X25
1.50-1.74	1X5Ø
1.75 - 1.99	1X75
2 - 2.24	2XØØ
2.25 - 2.49	2X25
3.25 - 3.49	3X25
8 - 8.24	8XØØ

RULES FOR CODING CHEMICALS

It has been necessary to formulate procedures for assigning descriptors in special subject areas.

For chemicals the following rules are in effect:

a. Chemical terms are identified as used in the report. "Hackh's Chemical Dictionary"(3rd ed.), Rev. and Edited by Julius Grant is used as the authority for chemical terminology. If a chemical is identified

by more than one term, the descriptor is the term most commonly accepted. Cross-references are used from the lesser known term.

b. The name of the main chemical compound in addition to the combined term is used. For Dibutyl the descriptors "Dibutyl" and "Butyl" are assigned so that the report may be retrieved by searching either descriptor.

c. For explosives the descriptors which constitute the explosive are used:

Trinitroethyl code Trinitro ethyl (two words)
Trinitrobutyrate code Trinitro butyrate (two words)

d. For explosives which are popularly referred to by abbreviations, the descriptor for abbreviation as well as the chemicals forming the explosive are used:

RDX code also Cyclo trimethylene trinitramine
PBX code also Plastic-bonded explosives

e. Where the abbreviation for a chemical is not in common usage, descriptors are assigned for the chemical only:

DATB code Diamino trinitro benzene
PETN code Penta erythritol tetra nitrate

TRADE NAMES

Trade names have descriptor codes established also. The materials or substances constituting the product require descriptors. "Chemical Trade Names and Commercial Synonyms" by Williams Haynes is used as the authority for trade names:

Mylar (Ethylene glycol terepthalic acid) (Plastic)
Teflon (Tetra fluoro ethylene polymer) (Plastic)
Lucite (Polymerized methyl methacrylate thermoplastic resins)
Cellosolve (Ethylene glycol monoethyl ether)
Amerite (Phenolic urea resins)

GEOGRAPHIC NAMES

Descriptors for geographic names are used where the location is an integral part of the report. Tests performed in the waters around Fort Lauderdale will yield different results from tests undertaken in the Bay of Fundy or the Barents Sea. In each instance, the geographic location where the tests were performed should be included.

SHIP NAMES

Descriptors for names of ships are coded, also the type of ships:

USS George Washington (Nuclear submarine)
HMCS Bonaventure (Aircraft carrier)
American Mariner (Maritime training ship)

PERSONAL NAMES

Names of individuals are listed where the person has given his name to a theory or widely accepted physical law:

Peltier (Solid state physics)
Fermi (Nuclear physics)
Hamilton (Mathematics)
Reynolds (Aerodynamics)

USE OF EQUIPMENT

Reports covering Sonar, Radar, Radio, Infrared, Radiation require the descriptor "Equipment". The exception is a theoretical study in these fields which discusses the laws governing the subject without a discussion of the equipment.

Large pieces of ordnance or equipment (missiles, reactors, satellites etc.) which do not carry a Mark designation but use numbers for various models or modifications have the number designated as part of the code term using five or six characters:

BULLPUP 7	BULL7
SNAP 10	SNPZ1Ø
TRANSIT 3	TRST3

ROLE INDICATORS

Descriptor codes utilising five letters have been established where the fifth letter serves as role indicator. Thus the letter P in the fifth position of a code refers to Properties, D in the fifth position, Design.

Lead Properties	LEADP
Mines Design	MINED

Role indicators are used with code terms when the report being analysed discusses more than one item, be it metal, chemical or a piece of ordnance. A descriptor could refer to more than one item and the role indicator serves to eliminate erroneous information or "false drops" in machine searches.

A quality control report is concerned with the weapon connector for the Mark 90 depth bomb. The descriptor "quality control" can be retrieved as quality control of the connector or quality control of the depth bomb. By the use of the letter Q for quality control as role indicator associated with the connector, a "false drop" in this area is eliminated:

Connectors quality control	CONRQ

The following role indicators have been established:

A	Agent (performing action)
B	Bibliography (when required)
C	Characteristics
D	Design, Description
E	Effects
F	Fabrication, Production
I	Instruction, Operation
O	Object (acted upon by agent)
P	Properties
Q	Quality control
R	Research
S	Study, State of the Art
T	Test, Testing procedures
V	Evaluation
X	Negative
Z	Performance, Behavior

NEGATIVE CONCEPT

By means of fifth letter role indicator, a negative concept may be coded.

Angle of attack	ANGE
Angle of attack (negative)	ANGEX
Persistent (of gases)	PSIS
Non-persistent (of gases)	PSISX
Wings	WING
Wingless	WINGX
Ice	ICEA
Ice free	ICEAX

APPENDIX B

EXAMPLE OF COMPLETED INPUT DATA WORKSHEET AND RESULTANT PUNCHED CARDS

(Extracted From NOLTR 62-51)

Changes since the data of publication of NOLTR 62-51 are not reflected.

APPENDIX B

EXAMPLE OF COMPLETED INPUT DATA WORKSHEET AND RESULTANT PUNCHED CARDS

IBM PUNCHED CARDS

IBM cards are punched from the information listed on the completed work sheet.

SOURCE	REPORT NUMBER	REPORT DATE	CLASS AND CODE	CATALOGER-DATE
TUIA	118Ø72	1159	UØ14	hcy 4-11-61

DESCRIPTORS	CODES	DESCRIPTORS	CODES
Explosions	EXPL	Equipment	EQUI
Implosions	IMPO	Procedure	PROC
Spheres	SPHE		
Air	AIRE		
Blast	BLAS		
Gas	GASE		
Dynamics	DYNA		
Point source	PONT		
Wave	WAVE		
Shock	SHOC		
Effects	EFFE		
Experiment	EXPE		

IBM WORK SHEET - LIBRARY DIVISION [illegible]

Figure 2 Completed IBM Work Sheet

IBM cards are punched on the 026 key punch, utilizing the first 72 columns for codes. The first 36 columns are reserved for identification information, the remaining 36 (columns 37-72) are used for punching subject information by use of descriptor codes.

Columns 1-6 are used for source code. If four columns only are required, columns 1-4 are punched, with columns 5 and 6 unpunched.

Columns 7-12 are used for report number (or series for NOL-originated publications). All six columns are used for report number. If the report number has less than six digits, zeros precede the number to make a total of six.

Columns 13-16 contain codes for the date of the report, using four columns for the code.

Columns 17-18 are not punched.

Columns 19-22 are used for security classification of the report and total number of descriptor codes assigned. Four columns must be used for this information. Security classification is represented by a single letter and three numbers are required for the number of descriptor codes. If the total number of descriptor codes is less than three digits, zeros precede the number of descriptor codes.

Columns 23-36 are used for circulation limitations and/or bibliographic information. If one circulation limitation is required, columns 23-26 are used. If an additional circulation limitation is required, columns 27-30 are used; a third circulation limitation is punched in columns 31-34. Circulation limitation is always restricted to four characters.

Bibliographic codes may be four to six characters. If the identification requires one bibliographic code, columns 31-36 are used. This may be a four character code (31-34), five character code (31-35) or a six character code (31-36). If a second bibliographic code is necessary, columns 27-30 are used. Second and third bibliographic codes may not exceed four characters. The third bibliographic code is punched in columns 23-26.

Descriptor code terms are punched in columns 37-72. Thus the first descriptor code is punched in columns 37-42. A four character code is punched in columns 37-40, a five character code in columns 37-41, a six character code in columns 37-42. Punching always begins at the left with unpunched columns at the right. The second descriptor code is punched in columns 43-48, the third columns 49-54, fourth in columns 55-60, fifth in columns 61-66, sixth in columns 67-72.

There cannot be more than six descriptor codes in columns 37-72. If more than six descriptor codes are required for a report, a second IBM card is punched. By means of a drum card, the identification information punched in the first thirty-six columns is repeated on all subsequent cards associated with a report. On the second card, additional descriptor codes are punched commencing with column 37. Seven to 12 descriptor codes will require two IBM cards, 13 to 18 codes three IBM cards, 19 to 24 codes four IBM cards, etc.

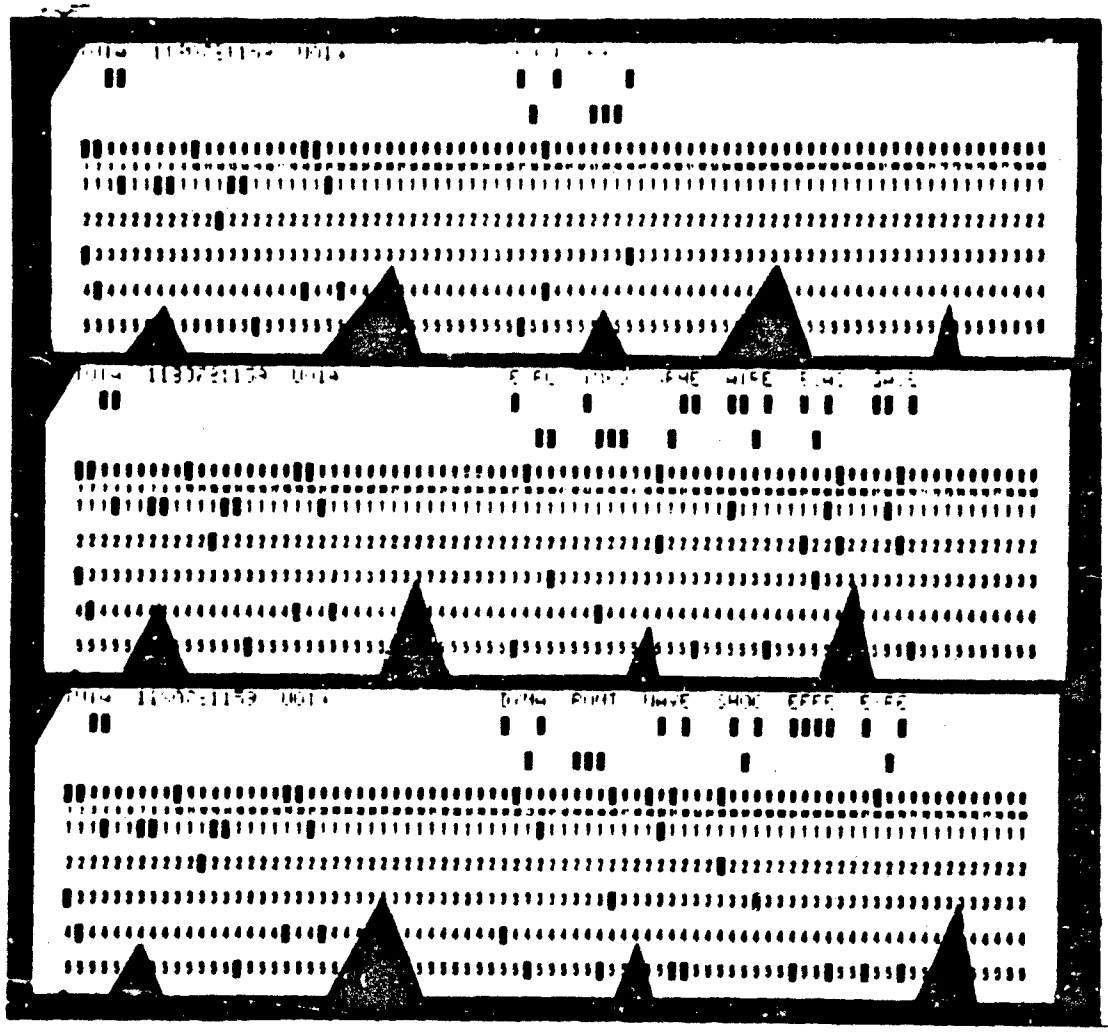


Figure 3 Code Input Information

APPENDIX C

SAMPLE SEARCH OUTPUTS

IDENTIFICATION			LIMITATIONS		
REQUEST	REQUESTOR	STAFF MEMBER	DATE	SECURITY	CIRCULATION
660042	0820	EL	L) 00 U) 00	L) U U) U	

CODE WORDS VACM SNAA

0001 DOCUMENT FOUND

AGENCY	ACCESSION NUMBER	SECURITY LEVEL	DATE	IMPORTANT CODE WORDS
DCFL	167180	U	07 65	ELEC LEAK OSCL VACM

IDENTIFICATIC			LIMITATIONS		
REQUEST	REQUESTC	PAPER	DATE	SECURITY	CIRCULATION
660042	0920		LI 00 UI 00	LI U UI U	

CODE WORDS VACM

0001 DOCUMENTS FOUND

U-16718?
HARRY DIAMOND LABORATORIES.
ELECTRICAL LEAKAGE IN THE SN1951B, BY HAROLD EDWIN
BOESCH, JR. JULY 1965.
4 P. DIAGRS. HDL TM-65-34 .

APPENDIX C(2)

APPENDIX D

CATALOG CARDS

A conventional library card catalog is maintained for ready reference. The descriptive cataloging of the reports, via the Friden Programmatic Flexowriter, produces a six-part, card file format. Five of these are filed in the card catalog under the following categories: source, document number, title, series. A sample catalog card is shown below. This card was produced at the same time that the charge card shown in Appendix E was produced. Permanent cards are produced once a week by the NOL Photographic Division when the Accessions List is printed.

PLASTICS

DATE	SUBJECT CATALOGER	REPORT NO
		U-168123

OC n

Picatinny Arsenal, Dover, N.J.
Dilute solution properties of high-molecular-wight finding molecular. Oct. 1965.
v.p.

templ0/28/65Jg

PAPER PATENTED BY NCR COMPANY

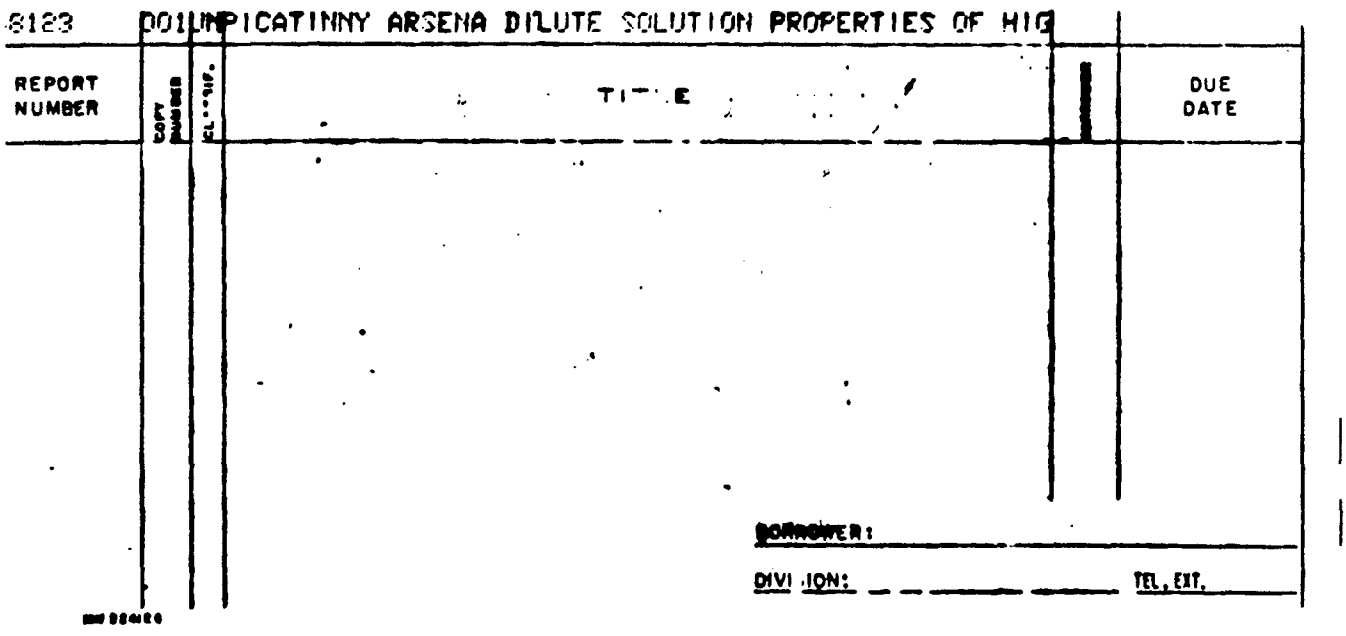
REPORT CATALOGING-PRNC-NOL-5070-7 (8-60)

APPENDIX E

CHARGE CARDS

For each report folder kept by the Library, a charge card is produced (NOL Form PRNC-NOL-5070/4 Report Loan Record) and filed with the report. The charge card is prepunched with accession number, copy number, classification, and title.

When a report is charged out, the borrower's code--the last four digits of his payroll number--and the date of charge are entered on the card. The charge cards are automatically produced on an IBM 026 key-punch tied to a Friden Programmatic Flexowriter. A sample charge card is shown below.



APPENDIX F

ACCESSIONS LIST

At the same time the descriptive cataloging is being accomplished, the Flexowriter produces a paper tape with the same information. These tapes are then filed first by subject and then under each subject by daily sequence (which results in filing by accession number).

At the end of each week, the paper tapes are read by the Flexowriter which automatically types the accessions lists on paper mats for reproduction. The mats are checked for errors and then sent to the NOL Photographic Division which produces the accessions list and permanent catalog cards. An example of entries on the accessions list follows:

U-164297

Sperry Gyroscope Co., Syosset, N.Y.
Multi-speed repeater Mk 3 mods 0,1,2,3. 1st. rev. June 1965.
v.p. charts, tables. (NavShips 324-0535).
Supersedes NavShips 324-0535-, dated June 1962 and addendum dated April 1963.

I. Title. II. Series.

APPENDIX G

DEVELOPMENT OF DESCRIPTORS AND COMPUTER CODES

(Extracted From NOLTR 64-20)

Changes since the date of publication of
NOLTR 64-20 are not reflected.

APPENDIX G

DEVELOPMENT OF DESCRIPTORS AND COMPUTER CODES

Descriptors and computer codes are used by the NOL Library in its information storage and retrieval program. A preliminary list of 2500 descriptors was generated based on the subject headings and corporate authors used in the Library card catalog. Additions, changes, and deletions have been made as the number of technical reports coded for retrieval increased and as searches revealed need for changes.

There are now 22,000 descriptor codes in the Library computer program. Of this number, 7,500 descriptors are used for subjects. The remainder of the descriptors cover equipment, trade names, projects, geographic locations, corporate authors, personal names, etc.

The original plan to retain single terms whenever possible has been changed as the need arose for bound terms to permit more precise coding of reports and to eliminate false retrieval of information.

An alphabetical file of descriptors and computer codes is maintained by the Library on IBM cards. A second file, arranged alphabetically by computer code, is also retained on IBM cards. The alphabetical codes are maintained on magnetic tape, and this provides the IBM 7090 dictionary for the Library computer program.

Magnetic tape is updated weekly to permit incorporation of new codes and the addition of coded information on incoming technical reports. A major update of the entire Library tape containing report information is done annually. At this time, the current file of computer codes transferred to magnetic tape forms the basis for a new dictionary tape. Deleted codes are removed manually from descriptor and code file, and are therefore not available on the new tape. The revised magnetic tape contains only those codes which are used for analyzing and coding technical reports.

A table of four-letter alphabetical combinations was constructed to facilitate assignment of computer codes and to avoid duplication of codes. Table of Four-Letter Computer Codes Used in Library Retrieval Program, by Eva Liberman and Joyt L. Stevens, has been issued as NOLTR 62-50.

Descriptors for nouns are given in plural form with the exception of chemicals. Chemicals are in singular form to facilitate filing arrangement of chemical compounds.

Code and code also references are used in the listing.

Definitions used in identifications are brief, limited largely by the information which can be punched on an 80-column IBM card. A number of standard technical reference sources were consulted in identifying descriptors.

The mnemonic feature of selecting codes for a descriptor whenever possible assists the coder in assigning codes and also assists the searcher in his request for a computer search. For example, mines, MINE; missile, MISL; Miss-distance (Indicator), MISD; Aerojet General Corporation, AERJ; Naval Ordnance Laboratory, White Oak, NOLA; Naval Ordnance Test Station, Inyokern, California, NOTS.

All number codes are omitted from the descriptor list since procedures for coding numbers have been established. Numbers having less than four characters are preceded by zeros. For example, 1 is coded 0001; 100 is coded 0100. Numbers having four to six characters are coded as follows: 1,000 is coded 1000; 100,000 is coded 100000. Numbers having more than six characters use coded descriptors such as one million, two million, etc.

An identification following a descriptor term indicates to the coder that the general term is also coded. Thus, Paper-lined (Batteries) restricts the descriptor Paper-lined to Batteries and requires codes for Paper-lined and Batteries.

Trade names listed under Equipment Designations, Acronyms, Trade Names, etc., are identified as trade names where the descriptor could be so identified from standard reference sources. Where this fact cannot be ascertained, the descriptor is not identified as a trade name.

Government agencies, industries, and corporate authors which have had name changes retain a single code for all forms of the name. For example, Naval Propellant Plant and Naval Powder Factory have the code NPPL; Martin-Marietta Company, Martin Company, and Glenn L. Martin Company have the same code, MRTI. Army Electronics Research and Development Laboratories and Army Signal Research and Development Laboratories both use the code ASRD.

When the name of an organization as a separate entity is widely known, it has a separate code. Scripps Institute of Oceanography has a code other than the one used for the University of California. Armour Research Foundation has a code distinct from Illinois Institute of Technology.

Laboratories - whether Government, university, or industry - are identified by separate code where NOL's interests are closely tied to the Laboratory. If not, the Laboratory has the same code as the parent agent. Thus, Ballistics Research Laboratories, Aberdeen Proving Ground, is identified by BRLA; Aberdeen Proving Ground and other laboratories at Aberdeen Proving Ground are identified by APGA. Hanford Laboratories, operated by General Electric Company at Hanford, Washington, has its own code, as does Ordnance Research Laboratory at Pennsylvania State University. The field stations operated by National Aeronautics and Space Administration have the code NASA; however, separate codes are assigned to Goddard Space Flight Center, Greenbelt, Maryland; Manned Spacecraft Center, Houston, Texas; and George C. Marshall Space Flight Center, Huntsville, Alabama.

Individuals who have given their names to widely known theories are in the Corporate Author list rather than under subject. This is purely an arbitrary decision. Hall, Reynolds, Mach, von Karman, and many others are included in the Corporate Author list.

Acronyms are identified whenever possible within the limits of military security.

Role indicators are used with descriptor codes. This is a fifth letter following the four-letter computer code. The following role indicators have been established:

A - Agent (acts on an object)
B - Bibliography
C - Characteristics
D - Design, Description
E - Effects on, Effected by
F - Fabrication, Production
H - Specifications
I - Instructions, Operation
L - Development, History
O - Object (thing acted upon)
P - Properties
Q - Quality Control
R - Research
S - State-of-the-Art, Study
T - Tests
V - Evaluation
X - Negative (May be combined with any descriptor which does not have a separate code established for the negative)
Z - Performance

Examples: SUBMO, Submarine as object; AIRCA, Aircraft as agent; FUELP, Fuel properties; FUZED, Fuzes design; RADRI, Radar instruction or operation; TORPH, Torpedoes specifications; BATTQ, Batteries quality control; CNNEX, Unconnected.

Plans to use links to tie together related descriptors by means of a sixth letter in the computer code were abandoned due to limitations of the total number of codes which can be used.

The present computer program limits the total number of descriptor codes to 32,767 - the size of the memory of the 7090 computer.

Very few subject terms are now being added to the descriptor list. However, new codes for equipment, trade names, names of ships and submarines, and project names, are continually being added. When the total of 32,767 descriptors is reached, the Library program will require modification to permit additional descriptors or deletion of descriptors which are rarely searched.

APPENDIX H

SAMPLES OF DESCRIPTORS AND COMPUTER CODES

(Extracted From NOLTR 64-20)

Changes since the date of publication of
NOLTR 64-20 are not reflected.

APPENDIX H

SAMPLES OF DESCRIPTORS AND COMPUTER CODES

NOLTR 64-20

DESCRIPTORS AND COMPUTER CODES USED
IN NAVAL ORDNANCE LABORATORY RETRIEVAL PROGRAM

SUBJECTS (SINGLE AND BOUND TERMS)

DESCRIPTOR	CODE
ABBREVIATIONS	ABBR
ABERRATION (DEVIATION FROM NORMAL)	ABER
ABIETATE, ABIETIC ACID	ABIE
ABLATE, ABLATION (WEARING AWAY OF SURFACE MATERIAL)	ABLA
ABLATION METERS	ABLM
ABORT CODE FAILURE	
ABRASION, ABRASIVES	ABRA
ABSORBENT, ABSORPTION, ABSORPTIVITY	ABSR
ABSORBERS	ABSO
ABSORPTIOMETERS (MEASURES ABSORPTION OF GAS)	ABSM
ACACIA GUM CODE ARABIC GUM	
ACCELERATORS (INCREASES SPEED OF CHEMICAL REACTION)	ACCR
ACCELERATION, ACCELERATING (RATE OF CHANGE OF VELOCITY)	ACCL
ACCELERATION (NEGATIVE) CODE DECELERATION	
ACCELEROMETERS (MEASURES ACCELERATION)	ACCE
ACCEPTANCE	ACCT
ACCEPTORS	ACCP
ACCESS	ACSS
ACCESSORIES	ACCS
ACCIDENTS, ACCIDENTAL CODE INADVERTENT WHERE APPLICABLE	ACCI
ACCOUNTABILITY	ACCN
ACCUMULATION (COLLECT, BRING TOGETHER)	ACUM
ACCUMULATORS (DEVICE FOR STORING ELECTRICITY)	ACCU
ACCURACY (CORRECTNESS)	ACUR
ACETAL (DIETHYLACETAL)	ACES
ACETALDEHYDE, ACETIC ALDEHYDE CODE ETHYL ALCOHOL (TWO WORDS)	
ACETAMIDE (ACETIC ACIC AMIDE)	ACLD
ACETATE	ACTA
ACETIC ACID, ACETIC ESTER	ACET
ACETIN (GLYCEROL ACETIC ACID)	ACTI
ACETOL CODE ACETYL CARBINOL (TWO WORDS)	
ACETONATE	ACTN
ACETONE, ACETONYL (DIMETHYL KETONE)	ACTO
ACETOXY	ACTX
ACETYL, ACYL	ACTL
ACETYLENE (HYDROCARBON)	ACTY
ACETYLIDE	ACTY
ACID	ACID
ACONITE, ACONITUM (MEDICINAL PLANT)	ACNT
ACONITINE (ACETYL BENZOYL ACONINE)	ACON
ACOUSTIC FIELDS	ACOT

EQUIPMENT, ACRONYMS, TRADE NAMES, PROJECTS, CODE NAMES, ETC.

DESCRIPTOR	CODE
A (EQUIPMENT)	AAAA
A MARK (EQUIPMENT)	AMAR
AA CODE ANTI-AIRCRAFT, AIR-TO-AIR WHERE APPLICABLE	
AA (RADAR EQUIPMENT)	AAXX
AAA CODE ANTI-AIRCRAFT ARTILLERY WHERE APPLICABLE	
AAA (DETECTOR)	AAAX
AAD (DETECTOR)	AADX
AADS CODE ACTIVE ACOUSTIC DOPPLER SYSTEM	
AAFFR (ANTI-AIRCRAFT FOLDING FIN ROCKET)	AAFR
AAG (RADAR EQUIPMENT)	AAGX
AAM CODE AIR-TO-AIR MISSILE WHERE APPLICABLE	
AAQ (SPECTROMETER)	AAQX
AAR (RADIOMETER)	AARA
AARON WARD USS (DESTROYER)	AWAR
AAS (RADIOMETER)	AASX
AB (ANTENNA)	ABAA
AB (CHEMICAL WARFARE AGENT)	ABXX
AB1 (BIOLOGICAL AGENT) CODE BRUCELLA SUIS	
ABBOT (GUN)	ABBT
ABBOT USS (DESTROYER)	ABBD
ABF (AIRBURST FUZE)	ABFU
ABLE (AUTONETICS BASE LINE EQUIPMENT)	ABLQ
ABLE (PROJECT)	ABLE
ABLESTAR (MISSILE)	ABLS
ABN (FUZE)	ABNA
ABNER READ USS (DESTROYER)	ARED
ABO (IGNITER MIXTURE)	ABOZ
ABRAHAM LINCOLN USS (SUBMARINE)	ALIN
ABWR (BOILING WATER REACTOR)	ABWR
AC (ANTI-COUNTERMINE MECHANISM)	ACAA
AC (CHEMICAL WARFARE AGENT) CODE HYDROGEN CYANIDE (TWO WORDS)	
AC (ELECTRICITY) CODE ALTERNATING CURRENT	
AC (PROPELLANT)	ACPR
AC AND W (RADAR) CODE AIRCRAFT CONTROL AND WARNING (RADAR)	
ACANIA (INSTRUMENT TRAINING SHIP)	ACAN
ACCORDS (ELECTRONIC EQUIPMENT)	ACOR
ACM CODE ANTI-COUNTERMINE	
ACOR (PROJECT)	ACPP
ACR (GYROSCOPE)	AC[illegible]
ACRAWAX (SYNTHETIC WAX) (TRADE NAME)	ACR[illegible]
ACRILAN (METHYL METHACRYLATE) (TRADE NAME)	AC[illegible]
ACRYLOID (PLASTIC) (TRADE NAME)	A[illegible]
ACRYLON (ACRYLIC RUBBER) (TRADE NAME)	A[illegible]
ACU (ALTITUDE CONTROL UNIT)	AC[illegible]
ACUTE HMS (MINESWEEPER)	ACU[illegible]
AD (AIRCRAFT)	AD[illegible]
AD (EXPLOSIVE)	ADE[illegible]

CORPORATE AUTHORS, PERSONAL AUTHORS AND GEOGRAPHIC NAMES

DESCRIPTOR	CODE
ABBE (NAME)	ABBE
ABC CODE AMERICAN, BRITISH, CANADIAN	
ABC JOINT COORDINATING COMMITTEE	ABCJ
ABERDEEN PROVING GROUND, ABERDEEN, MD.	APGA
ABERDEEN PROVING GROUND, ABERDEEN, MD. BALLISTIC RESEARCH LABORATORY	BRLA
ABL CODE ALLEGANY BALLISTICS LABORATORY	
ACADEMY OF SCIENCES, USSR CODE AKADEMIIA NAUK SSSR	
ACF INDUSTRIES, INC.	ACFI
ACOUSTIC AND PRESSURE CHECK RANGE CODE NORFOLK NAVAL SHIPYARD	
ACOUSTICAL SOCIETY OF AMERICA, NEW YORK, N.Y.	ASAA
ADAMS (NAME)	ADAS
ADCOCK (NAME)	ADCC
ADELPHI COLLEGE, GARDEN CITY, N. Y.	ADEP
ADEN (COUNTRY)	ADNZ
ADEN (NAME)	ADEN
ADMIRAL CORP., CHICAGO, ILL.	AMIL
ADVANCE INDUSTRIES, INC., APPLETON, WIS.	AIIN
ADVANCED METALS RESEARCH CORP., SOMERVILLE, MASS.	ADVN
ADVANCED RESEARCH PROJECTS AGENCY	ARPA
ADVANCED SCIENTIFIC TECHNIQUES RESEARCH ASSOCIATES, INC.	ASTH
ADVISORY GROUP FOR AERONAUTICAL RESEARCH AND DEVELOPMENT (AGARD)	AGAR
ADVISORY GROUP ON ELECTRON DEVICES, ADVISORY GROUP ON ELECTRON PARTS	AGEP
AEC NUCLEAR CROSS SECTIONS ADVISORY GROUP	NCSA
AERCON, INC., PASADENA, CALIF.	ACOC
AERO GEO ASTRO CORP., ALEXANDRIA, VA.	AERG
AERO SERVICE CORP. DIV. LITTON INDUSTRIES, INC.	AERR
AEROCHEM RESEARCH LABORATORIES, INC., PRINCETON, N.J.	ACHE
AERODYNAMISCHE VERSUCHSANSTALT E.V., GOTTINGEN, GERMANY	AVGO
AEROFLOT (RUSSIAN AIRLINE)	AERF
AEROJET-GENERAL CORP., AZUSA, CALIF.	AERJ
AEROJET-GENERAL NUCLEONICS, SAN RAMON, CALIF.	AGNU
AERONAUTICAL CHART AND INFORMATION CENTER, ST. LOUIS, MO.	ACIC
AERONAUTICAL RADIO, INC. CODE ARINC RESEARCH CORP.	
AERONAUTICAL RESEARCH ASSOCIATES OF PRINCETON, INC., PRINCETON, N.J.	ARNR
AERONAUTICAL SYSTEMS DIV. CODE WRIGHT AIR DEVELOPMENT CENTER	
AERONCA MANUFACTURING CORP., MIDDLETOWN, OHIO	AERC
AERONUTRONIC, DIV. PHILCO CORP., NEWPORT BEACH, CALIF.	AERU
AEROPHYSICS DEVELOPMENT CORP.	ADCP
AEROPROJECTS, INC., WEST CHESTER, PA.	APIN
AEROSPACE CORP.	ARSP
AEROSPACE TECHNICAL INTELLIGENCE CENTER	ATIC
AEROVOX CORP., NEW BEDFORD, MASS.	ARVX
AFGHANISTAN	AFGH
AFRICA	AFRI
AGARD CODE ADVISORY GROUP FOR AERONAUTICAL RESEARCH AND DEVELOPMENT	

APPENDIX I

SPECIAL RECORD

(Extracted From NOLTR 64-125)

Changes since the date of publication of NOLTR 64-125 are not reflected.

APPENDIX I

SPECIAL RECORD

The special record appears as the first record of the Master dictionary tape. It contains counts and tables used in computing the zato codes and coding security and circulation limits.

Word No.	Definition
1	Descriptor Count
2	Document Count
3	Last random number generated for zato coding
4	$\sum_{i=1}^{n} u_i \log u_i$ when u_i = freq. ith descriptor/total no. documents
5	$\sum_{i=1}^{n} \bar{u}_i \log \bar{u}_i$ where $\bar{u}_i = 1 - (1-u_i)^6$
6	Last rank assigned to a descriptor
7	C (BCD,ØØØØØC)
8	D
9	R Table of Security Levels
10	S
11	T
12	U
13	777777777777
14-23	Ø
24	ACME (BCD,ØØACME)
25	ADCO
26	ANJM Table of Circulation Limits
27	FRDA
28	HODS
29	INFC
30	INRE
31	LIMI
32	NAAL

Word No.	Definition	
33	NACO	
34	NATO	
35	NOFO	
36	NRDD	
37	NRST	
38	OUOA	
39	PASO	
40	PPIN	
41	REDA	
42	SPCO	
43	TALI	
44	VTFU	
45	777777777777	
46-58	ø	
59	S	Octal Security Codes
60	11	
61	3	
62	7	
63	13	
64	1	
65-75	ø	
76	1	
77	2	Octal Circulation
78	3	Limit Codes
79	4	
80	5	
81	6	
82	7	
83	10_8	
84	11_8	
:		
96	25_8	
97-109	ø	

APPENDIX J

MASTER DICTIONARY RECORD

(Extracted From NOLTR 64-125)

Changes since the date of publication of NOLTR 64-125 are not reflected.

APPENDIX J

MASTER DICTIONARY RECORD

The following is a description of the master dictionary as it appears
ı core and on tape. Each record except for perhaps the last is 4200 words
ıng with 350 descriptors to the record and 12 words per descriptor.

ırd	Part	Definition
1		BCD Descriptor
2	DECR	Descriptor frequency
	ADDR	Descriptor rank
·10		8-word batch Zato Pattern
.–12		2-word Document Zato Pattern

APPENDIX K

FORMING THE ZATO CODES

(Extracted From NAVWEPS Report 7388 and NOLTR 64-125)

Changes since the date of publication of
NAVWEPS Report 7388 and NOLTR 64-125
are not reflected.

APPENDIX K

FORMING THE ZATO CODES

NAVWEPS REPORT 7388

The straightforward way to search a file which has not been arranged in any special order would be to examine it item by item and check for the presence of the requested descriptors among the twenty or so that have been assigned to each entry. Conversely, the criterion for rejection of an item is the absence of at least one of the requested descriptors. The chief drawback is that -- even in a collection of moderate size -- the probability that any given item will be relevant to the query is quite small; so that a large quantity of material must be carefully examined and then discarded (like looking for a needle in the proverbial haystack).

As a matter of fact, such an examination does constitute the final step of the program, but it is preceded by operations intended to eliminate much irrelevant material. To attain a high effective scanning rate these preliminary operations have been designed to separate extraneous material with as little effort as possible. The basic strategy adopted is that of repeatedly screening the recorded information of finer and finer levels of discrimination until only a small residue of relevant (or possible relevant) material is left. At each level the criteria for discrimination are more stringent but the quantity of material to be examined is smaller, so that the total effort expended remains reasonable.

To accomplish this task we need a means for sharply distinguishing between wanted and unwanted material, or -- in the language of the communication engineer -- what is required is a suitable filter. (In the case of the haystack, for example, a magnet would be an effective instrument for separating needles from hay.) For recorded information a good tool is the method of superposed random codes introduced by Calvin N. Mooers of the Zator Company (and sometimes referred to as Zatocoding). It is a scheme for shunting aside extraneous items on a statistical basis, and is readily adapted for use with digital computing machines.

The method is based on a comparison of data in a single field of binary digits. Most of the unwanted items can be rejected by this one test. However, all relevant items plus a few others are allowed to pass. The ones which do pass can be subjected to additional tests to cull out most of those which accidently get by previous stages of inspection.

The essential feature of the method is the assignment of a random pattern of marks and spaces to each descriptor or keyword in the dictionary. On a notched-edge card these are represented by notches and blanks. In the electronic computer they correspond to 1's and 0 s. The patterns consist mostly of spaces with only occasional marks, and are generated by a special computer routine. The number of marks placed in each pattern is adjusted to be proportional to the discriminating power of the corresponding term. Terms with a high frequency of occurrence offer little selectivity, so they are encoded with relatively few marks, and vice versa.

The pattern for the totality of descriptors which have been assigned to one particular document is formed by superposing the patterns for each individual descriptor. As far as each punching position or binary digit is concerned, we form what is referred to as the logical sum of the digits. That is, wherever a one (or notch) appears in any of the descriptor patterns, a one will appear in the sum. The only positions which will end up with zeros are those which were zero in all the individual patterns.

The query patterns are formed in a similar manner in response to specific requests for searching the tape file. However, queries are usually framed by selecting only a very small set of descriptors (from 1 to 7 in this system). Consequently, their patterns will contain relatively few 1's or marks.

During a search each of the document patterns stored on the magnetic tape is successively compared with the list of query patterns. The sole condition that an item should pass the screening test is that it should have a mark at every position where the query has a mark. If it cannot satisfy this condition, then the document cannot have among its assigned descriptors the particular set which form the query. In this case the item is immediately rejected as being not relevant. Note that a failure at a single position is sufficient to reject an item. On the other hand, there is no guarantee that those few that do pass will prove to be relevant, but in any case only this residue need be subjected to further testing.

The final secreening does consist of a comparison of individual query terms against document terms to weed out any extraneous items which by chance may have passed all previous levels of testing. Thus, the final result is a list of just those items which are pertinent to the stated query and no others.

To understand why the searching can proceed at such a rapid pace it is worth taking a moment to look at a small fragment of the actual program, specifically, at the innermost loop of the searching routine. By this is meant the most repetative portion of the program. This kernel contains only three instructions. The first one calls up one word (i.e., 36 bits) of the coded pattern for a document; the second instruction compares this with the corresponding coded pattern for one of the requests being processed; and the third instruction is a conditional transfer or branch. If the item is rejected by the query, this instruction sets up a similar comparison with the next query on the list; or, if it is not rejected, causes additional words of the same two patterns to be compared until either a rejection does occur or, in rare cases, until it is necessary to subject the item to an individual term-by-term inspection.

Calculations for the formulation of the zato codes follow.

Definitions

n = number of descriptors

f_i = the ith descriptors frequency

D = number of documents

b_i = number of "1" bits in the ith descriptor zato

Document Zato (2 word, 72 bits)

1. $u_i = f_i/D$

2. $L = 72$

3. $$K = \frac{L}{\sum_{i=1}^{n} u_i \log u_i}$$

4. $$b_i = K \log u_i = \frac{72 \log u_i}{\sum_{i=1}^{n} u_i \log u_i}$$

Batch Zato

1. $\bar{u}_i = [1 - (1 - u_i)^6]$

2. $\bar{L} = 288$

3. $$\bar{K} = \frac{\bar{L}}{\sum_{i=1}^{n} \bar{u}_i \log \bar{u}_i}$$

$$4.\quad \bar{b}_i = \bar{K} \log \bar{u}_i = \frac{288 \log \bar{u}_i}{\sum_{i=1}^{n} \bar{u}_i \log \bar{u}_i}$$

After the number of bits to appear in a zato pattern are computed, a random number generator is used to place the bits in their respective 2-word or 8-word array.

APPENDIX L

DOC ARRAY

(Extracted From NOLTR 64-125)

Changes since the date of publication of NOLTR 64-125 are not reflected.

APPENDIX L

DOC ARRAY

The DOC array is used in the file maintenance program to store the incoming documents in their coded form prior to being grouped in batches. The file maintenance program also uses it as the primary array where deletions and I. D. change information is stored. The following table will show the three ways in which the DOC array is used.

1. Temporary document storage

Word	Bits	Definition
1-8		Documents 8-word Batch zato
9-10		Documents 2-word document zato
11	S,1-19	Document number in binary
	20-24	1st circulation limit
	25-29	2nd circulation limit
	31-35	3rd circulation limit
12	S,1-17	Document descriptor count
	18-22	Month in binary
	23-29	Year of 20th century
	30-35	Security
13		1st word of Miscellaneous infor.
14		2nd word " " "
15		3rd word " " "
16		AAAABB
17		BBCCCC Four most important descriptors
18		DDDD
19		Issuing Agency

2. I. D. Changes

Word	Bit	Definition
1-10		Zero
11	S,1-19	Document number in binary
	20-24	1st circulation limit
	25-29	2nd " "
	31-35	3rd " "
12	S,1,17	Zero
	18-22	Month
	23-29	Year
	30-35	Security
13		1st word of Miscellaneous infor.
14		2nd word " "
15		3rd word " "
16		Issuing Agency
17	S,1-19	New document number
	20-24	New 1st circulation limit
	25-29	" 2nd " "
	31-35	" 3rd " "
18	S,1-17	Zero
	18-22	New Month
	23-29	New Year
	30-35	" Security
19		" 1st Miscellaneous infor.
20		" 2nd " "
21		" 3rd " "
22		" Issuing Agency

3. Deletions

Word	Bits	Definition
1-10		Zero
11	S,1-19	Document number
	20-24	1st Circulation limit
	25-29	2nd " "
	31-35	3rd " "
12	S,1-17	Zero
	18-22	Month
	23-29	Year
	30-35	Security
13		1st Miscellaneous infor.
14		2nd " "
15		3rd " "
16		Issuing Agency
17-22		Zero
23		Next document number
⋮		

APPENDIX M

PACKING THE DESCRIPTOR RANKS

(Extracted From NOLTR 64-125)

Changes since the date of publication of NOLTR 64-125 are not reflected.

APPENDIX M

PACKING THE DESCRIPTOR RANKS

After the ranks have been sorted and differenced Index term numbers are computed to replace the Δ ranks in the following manner:

Index term numbers consist of a four-bit stem and a suffix of no more than 14 bits.

Stem = Count of the number of bits to the right of the first non-zero bit in the Δ rank.

Suffix = The bits to the right of the first non-zero bit in the Δ rank.

Index term numbers vary in length between four and eighteen bits.

After the Index term numbers are computed for all Δ ranks, they are packed in the first 35 bits (S,1-34) of successive words with more than one term number per word. Bit 35 is zero for every word in which term numbers for a given document lie except for the last word in which case it is set equal to "1" thus terminating the sequence for a given document.

APPENDIX N

DOCUMENT FILE RECORD
(ALSO BATCH ARRAY)

(Extracted From NOLTR 64-125)

Changes since the date of publication of
of NOLTR 64-125 are not reflected.

APPENDIX N

BATCH ARRAY

The following is a description of the document file. The batches described below are as they appear in core and within the document records on tape. Each document record contains a variable number of batches and each batch has a variable number of words.

Word	Bits	Definition
1	3-17	Relative address of next batch
2-9		Batch ZATO = $\sum_{i=1}^{p}$ [Batch Zato of ith DOC in batch]
10-11		DOC 1's document ZATO
12-20		DOC 1's IDENTIFICATION
21-22		DOC 2's ZATO
23-31		DOC 2's IDENTIFICATION
32		
⋮		
43		
⋮		
54		
⋮		
65-66		DOC 6's ZATO
67-75		DOC 6's IDENTIFICATION
76		DOC 1's Index term Nos.
⋮		⋮ variable length
⋮		DOC 2's Index Term Nos.
⋮		

The following table will concern itself with the individual document within the batch:

Word	Bits	Definition
10		First word of DOC 1's zato
11		Second word of " "
12	S,1-19	Document No. in binary
	20-24	Circulation limit No. 3
	25-29	Circulation limit No. 2
	31-35	Circulation limit No. 1
13	S,1-17	Address of DOC's index term number relative to beginning of batch
	18-22	Month in binary (18 in month means 10th Century)
	23-29	Year of 20th Century
	30-35	Security
14		First word of Miscellaneous infor.
15		Second word of " " B.C.D.
16		Third word " " "
17		AAAABB Four most important code words
18		BBCCCC
19		DDDDDD
20		Issuing Agency in B.C.D.

APPENDIX O

TITLE FILE RECORD
A AND B ARRAYS (STACKED TITLE TAPE)
DELET ARRAY

1. TITLE FILE RECORD (ALSO TF ARRAY)

The following is a description of the title file as it appears on tape and in core (TF array). Each record is of variable length and is limited to a maximum of 500 words.

First Word—Document number; count of number of words in the document title; relative address of next title.

Second Word—Used to indicate if the document is an NOL publication. If it is, the word will contain the code number of the NOL publication. If it is not, it will be blank.

Third to n^{th} Word—Bibliographic entry (string of characters).

2. A AND B ARRAYS

The following is a description of the A and B arrays as they appear in core. The A and B arrays are used as temporary storage of the new titles to be added to the master title tape. The tape containing the new

titles is referred to as The Stacked Title Tape. Each record is of variable length and is limited to a maximum of 200 words. There is one title entry per record.

First Word—Special word containing word count of record in decrement. It is used as a control word in the generalized sort program.

Second Word—Document Number.

Third Word—Code for NOL publication or blank.

Fourth to n^{th} Word—Bibliographic entry (string of characters forming the structured format of the entry).

3. DELET ARRAY

The DELET array consists of two words per title entry to be deleted.

First Word—Document Number.

Second Word—Code for NOL publication or blank.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

ORIGINATING ACTIVITY (Corporate author) BOOZ - ALLEN APPLIED RESEARCH, INC. 4733 Bethesda Avenue Bethesda, Maryland 20014	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Mechanization Study of the Technical Library, Naval Ordnance Laboratory, White Oak, Maryland

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report of on-site survey

5. AUTHOR(S) (Last name, first name, initial)

G. A. Kershaw, D. Crowder, J. E. Davis, E. G. Loges, E. Merendini, S. M. Thomas

6. REPORT DATE September, 1966	7a. TOTAL NO. OF PAGES 117	7b. NO. OF REFS 5
8a. CONTRACT OR GRANT NO. DSA-7-15489 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 914-1-27 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AD 640 122	

10. AVAILABILITY LIMITATION NOTICES

Distribution of this Document is unlimited

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Defense Supply Agency Defense Documentation Center Cameron Station, Virginia

13. ABSTRACT

The Library at the Naval Ordnance Laboratory (NOL) uses the IBM 7090 computer for information storage and retrieval of its technical reports collection. All technical reports are eventually entered on microfilm, with the present collection totaling about 150,000. However, only documents received from 1958 to the present are available for computer searches. Plans call for an annual listing by EAM techniques, of periodicals received at NOL. During fiscal 1966, punched card preparation will be completed. In addition, NOL and the Applied Physics Laboratory at Johns Hopkins are considering the possibility of cooperatively producing an annual listing of periodicals available in technical libraries in the Washington Maryland area. There is also a study being conducted at present to determine the feasibility o a printed book catalog. The computer program has provided more exhaustive searches and more detailed subject analysis of reports and has eliminated the time lag in processing subject information.

DD FORM 1473

Unclassified

Security Classification

Security Classification

KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Digital Computer						
Electronic Accounting Machines						
Information Retrieval						
Libraries						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

Security Classification